environmental microbiology

Environmental Microbiology (2016) **18**(8), 2721–2731 doi:10.1111/1462-2920.13299

Dynamic, mechanistic, molecular-level modelling of cyanobacteria: *Anabaena* and nitrogen interaction

Ferdi L. Hellweger,[1]* Neil D. Fredrick,[1] Mark J. McCarthy,[2,3] Wayne S. Gardner,[2] Steven W. Wilhelm[4] and Hans W. Paerl[5]

[1]*Department of Civil and Environmental Engineering, Northeastern University, Boston, MA, USA.*

[2]*Marine Science Institute, The University of Texas at Austin, Port Aransas, TX, USA.*

[3]*Department of Earth and Environmental Sciences, Wright State University, Dayton, OH, USA.*

[4]*Department of Microbiology, University of Tennessee, Knoxville, TN, USA.*

[5]*Institute of Marine Sciences, The University of North Carolina at Chapel Hill, Morehead City, NC, USA.*

Summary

Phytoplankton (eutrophication, biogeochemical) models are important tools for ecosystem research and management, but they generally have not been updated to include modern biology. Here, we present a dynamic, mechanistic, molecular-level (i.e. gene, transcript, protein, metabolite) model of *Anabaena* – nitrogen interaction. The model was developed using the pattern-oriented approach to model definition and parameterization of complex agent-based models. It simulates individual filaments, each with individual cells, each with genes that are expressed to yield transcripts and proteins. Cells metabolize various forms of N, grow and divide, and differentiate heterocysts when fixed N is depleted. The model is informed by observations from 269 laboratory experiments from 55 papers published from 1942 to 2014. Within this database, we identified 331 emerging patterns, and, excluding inconsistencies in observations, the model reproduces 94% of them. To explore a practical application, we used the model to simulate nutrient reduction scenarios for a hypothetical lake. For a 50% N only loading reduction, the model predicts that N fixation increases, but this fixed N does not compensate for the loading reduction, and the chlorophyll *a* concentration decreases substantially (by 33%). When N is reduced along with P, the model predicts an additional 8% reduction (compared to P only).

Received 18 December, 2015; revised 12 March, 2016; accepted 12 March, 2016. *For Correspondence. E-mail ferdi@coe.neu.edu; Tel. (617) 373-3992; Fax (617) 373-4419.

Introduction

Models are essential tools in microbial ecology, and their importance is growing with the volume of data and our mechanistic understanding and appreciation of the complexity of microbial systems. Dynamic phytoplankton (eutrophication, water quality, biogeochemical) models were originally developed in the 1970s (Chapra, 1997), and they typically simulate phytoplankton as 'concentration' state variables using differential equations. Growth is considered a 'reaction' between phytoplankton and nutrient molecules using simple Monod-type kinetics. At the time of their development, the models were compatible with the available observations (i.e. bulk measurements, such as chlorophyll *a*) and our mechanistic understanding of phytoplankton. However, observational technologies have advanced rapidly and provide data at the molecular level, including gene (Oh *et al.*, 2011), transcript (Vila-Costa *et al.*, 2013), protein (Hanson *et al.*, 2014) and metabolite (Ankrah *et al.*, 2014) levels, as well as metabolic activity for single cells (Musat *et al.*, 2008) and entire communities (Steffen *et al.*, 2014b). Also, our mechanistic understanding of microbes has increased tremendously since the models were originally developed. As a community, we have sequenced the genomes of many microbes and mapped their complete intracellular metabolic network (Nogales *et al.*, 2012). Many mechanisms, such as phenotypic differentiation (e.g. N-fixing heterocysts) are understood at the molecular level (Flores and Herrero, 2010; Kumar *et al.*, 2010; Muro-Pastor and Hess, 2012). Development of models has been slower, leading to an increasing disconnect between observations and models (Fuhrman *et al.*, 2013; Kreft *et al.*, 2013; Hellweger, 2015), which now limits the utility of models to support research and management of lakes, especially nutrient-impacted, cyanobacteria-dominated lakes like Lake Erie (US-Canada) and Taihu (China) (Paerl *et al.*, 2011; Steffen *et al.*, 2014a). The optimal level of detail and complexity in a model depends on many factors, including the purpose

of the model and data availability (Grimm *et al.*, 2005). However, it is clear that models presently lag behind observations, and that incorporating modern biology has the potential to improve them substantially.

Our long-term goal is to develop an ecosystem model based on modern biology and use it to support limnetic research and management. Here, we present one essential component of such a model, a mechanistic, molecular-level model of *Anabaena* – nitrogen (N) interaction. *Anabaena* is an ecologically important (bloom-forming, toxin-producing) cyanobacterial genus that can fix N2 gas. From a modelling perspective, this is an interesting problem, because the process is incompatible with a major product of photosynthesis (oxygen). To overcome this, *Anabaena* has evolved to differentiate specialized cells (heterocysts), which are internally anoxic and fix N, and then pass this fixed N on to adjacent vegetative cells in the filament. Much is known about the mechanisms underlying heterocyst development and N interaction in *Anabaena* (Flores and Herrero, 2010; Kumar *et al.*, 2010; Muro-Pastor and Hess, 2012), which makes this model development feasible.

The purpose of the model is to reproduce the observed complex behaviour of *Anabaena* – N interaction. For model development, we followed the pattern-oriented modelling (POM) approach, a general framework for designing models of complex systems (Grimm *et al.*, 2005; Hellweger, 2010; Koleva and Hellweger, 2015). In this case, the process started with a conceptual diagram, including known components and structure of the N assimilation system in *Anabaena* [fig. 1 in ref. (Flores and Herrero, 1994)]. A literature review identified a number of observed emerging patterns that characterize relevant behaviour of *Anabaena* (e.g. an increase of nitrogenase activity when fixed N is removed). Then, the model structure (e.g. the form of an enzyme rate equation) and parameterization were modified to reproduce the observed patterns. This was done in a manual and iterative manner within the constraints of known mechanisms and parameter values in the literature (where available). At each step, the current version of the model is calibrated as well as possible to the observations. When there are patterns the model cannot reproduce, the model structure is modified, and the process is repeated. The result is a structurally realistic model, which produces emerging behaviour consistent with observations.

We present the model of N assimilation by *Anabaena*, the database used for model development and the model-data comparison. This exercise shows that the model reproduces the majority of observed patterns, and often helps interpret observations. The model is then applied to a hypothetical lake to explore the efficacy of N reduction to control eutrophication, and it predicts that reducing N loading can lead to substantial improvement in lake conditions.

Model description

The model builds on recent reviews of N interaction and heterocyst differentiation in *Anabaena* (Flores and Herrero, 2010; Kumar *et al.*, 2010; Muro-Pastor and Hess, 2012), previous molecular-level models of cyanobacteria (Hellweger, 2010), and models of *Anabaena* (Hellweger *et al.*, 2008) and phytoplankton N assimilation (Flynn *et al.*, 1997). An overview is presented in Fig. 1, and full details are documented in the Supporting Information (SI). Following previous molecular-level models, we use a lumped approach, where a selected number of genes, proteins and metabolite pools are considered (Castellanos *et al.*, 2004; Hellweger, 2013). For intracellular N, the model considers nitrate (NO3), nitrite (NO2), ammonium (NH4), urea (URE), glutamine (GLN), glutamate (GLU), labile N (LN) and structural N (SN). NO3 and NO2 are taken up via the nitrate transporter (NrtA). NO2 also diffuses across the membrane. In the cell, NO3 is reduced to NO2 by nitrate reductase (NarB), and NO2 is reduced to NH4 by nitrite reductase (NirA). NH4 is taken up via the ammonia transporter (Amt1). URE is taken up via the urea transporter (UrtA) and hydrolyzed to NH4 by urease (UreA). In heterocysts, dinitrogen gas (N2) is fixed and converted to NH4 by nitrogenase (NifH), which is also exuded. NH4 and GLU are converted to GLN by glutamine synthetase (GS, GlnA), and GLN is converted to GLU by glutamate synthase (GOGAT, GlsF). Glutamate dehydrogenase (GDH) converts NH4 and 2-oxoglutarate (2OG) to GLU. GLN and GLU are metabolized to LN and SN. LN is also degraded back to GLN and GLU, so it serves as an N storage pool (discussed further below). Besides these N species, the model tracks P quota (QP), cell biomass (m) and 2OG. Within each cell, the model solves differential mass balance equations for transcript, protein and metabolite levels, using numerical solution methods (see Supporting Information). Enzyme rates vary dynamically, based on a number of factors. For example, the NifH rate is proportional to the NifH enzyme level and modified using Monod-type substrate limitation terms for N2 and light, and product inhibited by intracellular NH4.

The N assimilation enzymes are controlled by a number of regulatory proteins. NtcA binds to 2OG, and that complex (NtcA:2OG) controls the expression of multiple genes. The formation of the NtcA:2OG complex is a direct function of the availability of N. For example, an increase in NH4 leads to an increase in GLN, which leads to an increase in the GOGAT rate and a decrease in 2OG and NtcA:2OG. GlnB (PII) also binds to 2OG and exerts post-translational control over NrtA. Heterocyst differentiation is controlled by HetR and PatS. HetR is activated when the cell is stressed for N (sensed via NtcA:2OG). PatS is an inhibitor, which diffuses from committed heterocysts to neighbouring cells to repress their heterocyst differentiation process, so that

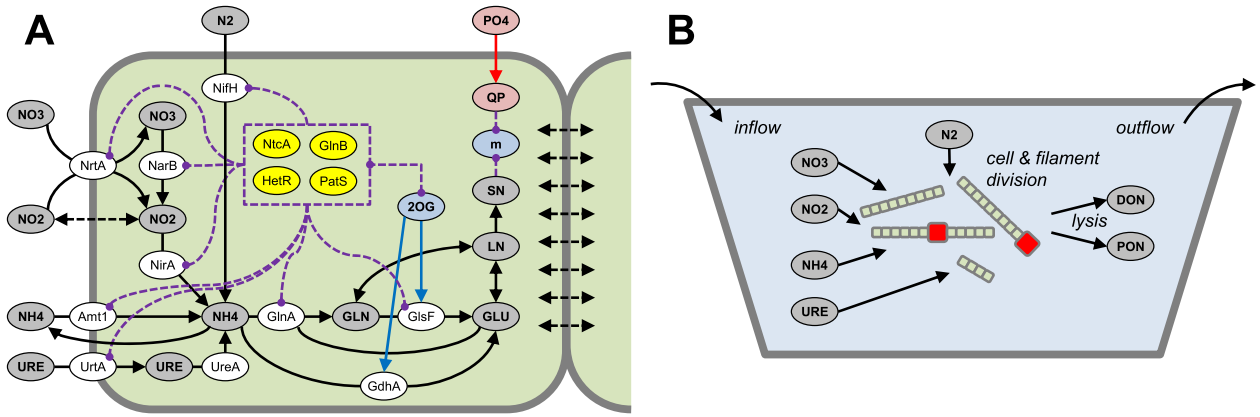

Fig. 1. Model overview.
A. Cellular processes. Simplified representation, see Supporting Information for full details. Blue, grey and red symbols correspond to C, N and P levels, white symbols are enzyme levels and yellow symbols are regulatory proteins.
B. Reactor/lake processes. Green and red cells are vegetative and heterocysts respectively. Grey symbols are extracellular concentrations. All symbols defined in text.

heterocysts form at about 10 cell intervals. Supporting Information Movie S1 follows a filament through NH4 to N-fixing conditions and illustrates the heterocyst differentiation process.

The model simulates individual cells that grow as a function of temperature, light and intracellular N and P content. They divide when they reach a threshold size, and can die and lyse. Cells form filaments that also divide (break up) once they reach a threshold cell number. Several intracellular compounds can diffuse along the filament. Figure 3C illustrates how heterocysts import GLU, combine it with fixed NH4 to form GLN and export GLN for a net export of N to vegetative cells. Filaments and cells are simulated as agents, and extracellular compounds are simulated as concentrations using an Eulerian approach. The model simulates various types of reactors, including batch, periodically-diluted batch, chemostat and turbidostat.

Some of the key agent-based modelling concepts used in the model are summarized here (see Supporting Information for details). Agents interact directly (e.g. via exchange of nutrients along the filament; Fig. 3C) and indirectly (e.g. competition for extracellular nutrients). They are simulated as super-individuals that represent a number of real individuals. A multi-level agent-based approach is used where a number of filament agents each include a number of cell agents, so a filament can be considered a 'collective'. The emergent properties include population size and behaviour (see Fig. 2 for examples). Individuals respond to their intracellular state and environmental conditions by changing gene expression, enzyme velocity, photosynthesis, etc. (see description of response to increasing NH4 above). The cells themselves can be considered complex systems and their individual behaviour an emergent property. Stochasticity is included in a number of processes, including cell death, lysis and division.

Database and model application

The model development builds on a database of observations, compiled from a literature review on *Anabaena* – N interaction. To keep the scope of the model development and data comparison manageable, we only used observations from laboratory monoculture experiments with wild-type cells under natural conditions. Therefore, field observations, observations from mesocosms or laboratory experiments with heterotrophs or other phytoplankton species or mutant *Anabaena* (those that are designed to behave differently) and/or metabolic inhibitors were not included.

Each experiment in the database was simulated using the model (i.e. repeated *in silico*). The model simulations generally followed the experimental protocols, which can be quite complex. For example, in one experiment (Ohmori *et al.*, 1977), cells were grown up in a high-nutrient medium for 3–4 days, put into nutrient-free medium (0.5–1 h), NO3 was added and then NH4 (after 10 min). In some instances, important experimental details are documented insufficiently. For example, although the light level is often specified, cell density and reactor vessel shape (affecting self-shading) are often omitted from experimental descriptions. In those cases, adjustments were made to the experimental protocols to obtain an optimum fit with the observations. For example, in some cases, we adjusted the reactor volume and depth to match the observations. Full details of the model applications are presented in the Supporting Information.

One model definition and parameterization is used for all experiments, so the model does not resolve differences

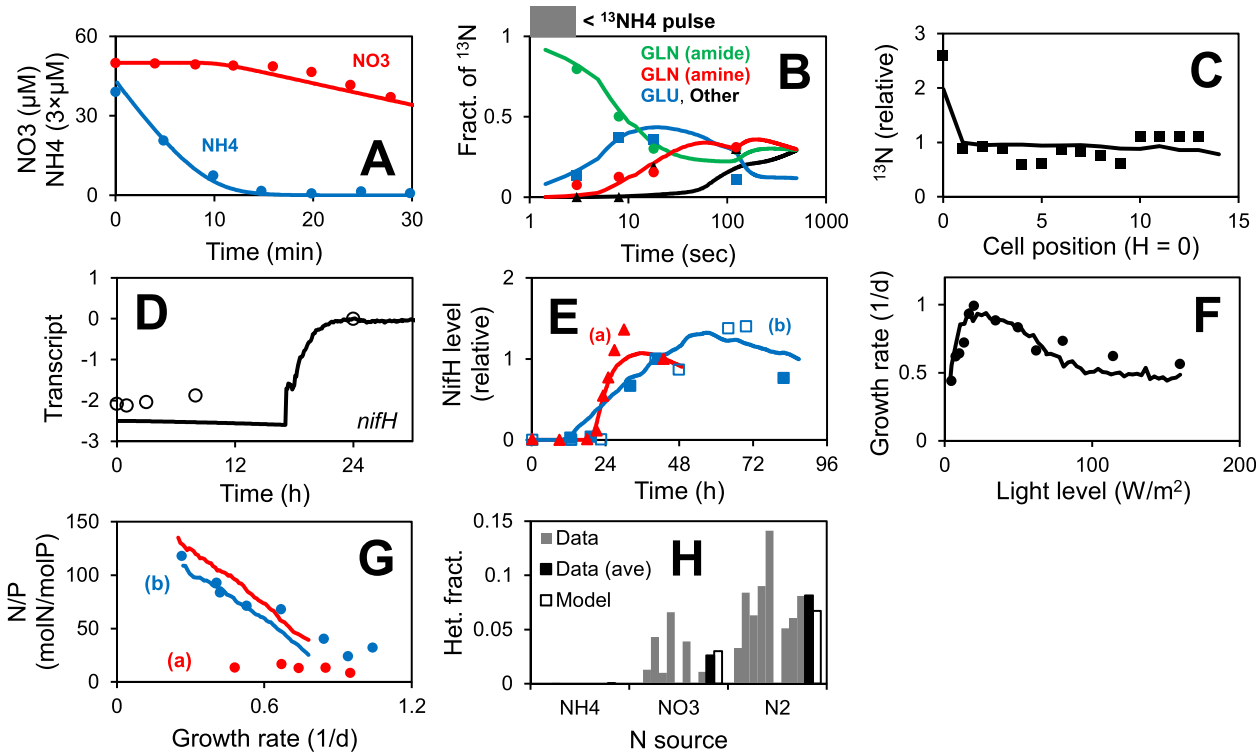

Fig. 2. Comparison of model to data. Representative examples. All comparisons presented in Supporting Information. Symbols are observed results from the literature and lines are model results (except for panel H, see legend).
A. Uptake of NO3 and NH4. Activation of NO3 uptake upon depletion of NH4.
B. Radioactive isotope tracing with ^{13}NH4. Pulse-chase experiment.
C. Radioactive isotope tracing with ^{13}N2. Labelling, 2 min. pulse. Radioactivity values normalized to average covered by data.
D. Changes in transcript levels following combined N removal. Values are relative, $\log_{10}$ of ratio to 24 h.
E. NifH level (enzyme activity, i.e. acetylene reduction) following combined N removal. NifH level is normalized to ~40h.
F. Turbidostat culture with N2 as N source at different light intensities. Growth rate vs. light level.
G. N/P ratio for cells grown in P-limited chemostat without fixed N. (H) Heterocyst fraction for growth on various N sources.
Data sources: (A) (Ohmori *et al.*, 1977), (B) (Meeks *et al.*, 1977), (C) (Wolk *et al.*, 1974), (D) (Ehira *et al.*, 2003), (E) (a) (Rowell *et al.*, 1977) (b) (Orr and Haselkorn, 1982), (F) (Lee and Rhee, 1999), (G) (a) (Rhee and Lederman, 1983) (b) (Layzell *et al.*, 1985), (H) Various, see Supporting Information.

between *Anabaena* strains. This approach necessitated making an adjustment to the temperature in some cases. For example, the model strain has an optimum temperature of 25°C and does not grow well at 30°C, which is the temperature used for most experiments with *Anabaena* sp. PCC 7120. Therefore, the temperature in those simulations was adjusted to 25°C to allow the model strain to grow. Also, some of the discrepancy between the model and data is possibly due to differences in strains. The model can resolve differences between strains by using different parameter values (e.g. higher optimal temperature for PCC 7120), and future laboratory or field applications may necessitate doing this.

The model definition and parameterization followed the pattern-oriented modelling (POM) approach, which involved explicitly identifying patterns in the observations and determining if the model reproduces the patterns. Patterns were defined as a relative change in some quantity, and model performance was judged in a binary manner. For example, the observations in Fig. 2G(a) show that the N/P ratio does not change with growth rate, which is a pattern that is not reproduced by the model. The observations in Fig. 2C show that N assimilation is initially higher in heterocysts compared to vegetative cells, a pattern that is reproduced by the model. The model skill is quantified as the fraction of patterns it reproduces. See the Supporting Information (Section S2.1.e) for additional discussion on defining patterns and model performance metrics.

Results and discussion

Model-data comparison

To ensure that our model agents behave like real *Anabaena* cells, the model development built on relevant observations from the literature. The objective was to constrain all of the pools and processes in Fig. 1. Our

database is quite extensive, encompassing 269 experiments from 55 papers published from 1942 to 2014. A substantial amount of research has been done on N fixation by *Anabaena*, so there are many papers to help constrain that process. Our database does not include observations for intracellular NO3, NO2 and URE concentrations, so there is considerable uncertainty in those model predictions. On average, there are observations from eight papers available to constrain each pool or process (see Supporting Information Fig. S9).

The presentation and discussion of the full model-data comparison for all of the 55 papers are necessarily included in the Supporting Information (Section S2). Here, we present and discuss in some detail a potpourri of selected representative samples ($\sim$5% of the entire database) (Fig. 2), followed by an overview of the results for the entire database. When cells are incubated with NO3 and NH4, they take up NH4 (their preferred N source) first and only assimilate NO3 when NH4 is depleted (Ohmori *et al.*, 1977) (Fig. 2A). The model reproduces this pattern, although it initiates NO3 uptake somewhat prematurely (prior to complete NH4 depletion) and underestimates the observed rate slightly. When cells are exposed to a short pulse of ^{13}NH4, the radioactivity is observed first in GLN-amide, followed by GLU, GLN-amine and then other components (Meeks *et al.*, 1977) (Fig. 2B). The model reproduces this pattern, but it underestimates GLU metabolism (GLU $>$ other) between 20 and 100 s. For N-fixing cells incubated with ^{13}N2, the radioactivity is initially higher in heterocysts and relatively uniform in the vegetative cells (Wolk *et al.*, 1974) (Fig. 2C). The model reproduces this pattern, but underestimates the difference between the heterocysts and vegetative cells. When cells are grown on fixed N and then transferred to media without fixed N, they differentiate heterocysts, and *nifH* expression increases after about 24 h (Ehira *et al.*, 2003) (Fig. 2D). The model reproduces this pattern. The nitrogenase activity also increases after about 24 h, but a slight temporary overexpression is observed in two different papers (Rowell *et al.*, 1977; Orr and Haselkorn, 1982), a pattern that is reproduced by the model (Fig. 2E, see discussion below). The growth rate is a function of light level, with limitation and inhibition regions (Lee and Rhee, 1999) (Fig. 2F). The model reproduces this pattern. When cells are grown in P-limited chemostats without fixed N, the observed N/P ratio in two papers is different (Fig. 2G). One study observed a relatively constant N/P ratio with growth rate, suggesting that the cells only fix enough N to support the present growth rate (Rhee and Lederman, 1983). However, another study found that the N/P ratio decreases with growth rate, suggesting that cells fix more N2 than they need at the lower growth rates (Layzell *et al.*, 1985). The cause of the discrepancy in the observations is unclear, and the model only reproduces one of these patterns. A summary of observations from eight papers shows that heterocysts are generally absent for cells grown on NH4, present at a low level for NO3 and higher for N2 (Fig. 2H). The model agrees reasonably well with these observations.

These and the other experimental observations in the database (see Supporting Information) were instrumental in building an understanding of the *Anabaena* N assimilation system through the $\sim$75 year history of research on this subject. Here, they are used to inform the structure and parameterization of the model. For example, the results of the isotope tracing experiment in Fig. 2B established the GS/GOGAT cycle as a major component of the N assimilation system, which influenced the pools and processes included in the model (Fig. 1A). Also, the rate of decrease of ^{13}N in GLU put quantitative constraints on the size of that pool. The shape of the growth rate vs. light curve (Fig. 2F) controlled the form (Haldane function) and parameterization of the underlying photosynthesis vs. light equation. The observed preference of NH4 over NO3 and the rates of uptake (Fig. 2A) put constraints on the N control system (i.e. the parameter controlling activation of *nrtA* expression by NtcA:2OG). An additional and more severe constraint was added by the interdependence of the patterns. It is relatively straightforward to adjust the model parameters to reproduce patterns in isolation. However, reproducing multiple patterns using the same model and parameterization is more difficult and the ability of the model to do this is support for its 'structural realism' (Grimm *et al.*, 2005).

The agreement between the model and observations is not perfect, but the model reproduces most of the patterns shown in Fig. 2. For the entire database of 55 papers, we identified 331 patterns, and the model reproduces 82% of them. The majority of mismatches involve inconsistencies in the observations. One example is the N/P ratio in Fig. 2G, where two similar experiments produced different patterns. Another example is an observation of substantial NH4 exudation and accumulation in the medium for N2-fixing cells (Subramanian and Shanmugasundaram, 1986), which is not common for uninduced (i.e. vs. GS blocked) conditions (Ramos *et al.*, 1984). Note that the model does include exudation of NH4 by heterocysts, based on (Paerl, 1984), but the NH4 is quickly assimilated by the vegetative cells and does not reach appreciable extracellular concentrations. Several mismatches are for experiments that showed no heterocysts or related parameters (NifH level, N fixation rate) in the presence of NO3. These observations vary considerably (Fig. 2H). In the model, heterocyst differentiation can be prevented simply by adjusting a parameter, but then the model would not match the observations where heterocysts are formed. Many mismatches are related to the changes of transcript levels when fixed N is removed. These observations are

noisy and often show inconsistent patterns (except for *nifH*). For example, when fixed N is removed, one paper observed little change in expression of *glnA* and *gdhA*, and another found almost an order of magnitude up- and down-regulation respectively (Ehira and Ohmori, 2006; Flaherty *et al.*, 2011). We cannot explain these inconsistencies, but presumably they are due to differences in strain, growth or experimental conditions, which are not considered in the present version of the model. If mismatches with inconsistent observations are removed, the model reproduces 94% of the observed patterns. The remaining mismatches highlight shortcomings in the model. For example, one observation showed that the fraction of dead vegetative cells is higher for growth on N2 vs. NO3 (Lee and Rhee, 1997). The authors of that paper suggest some mechanisms, such as the availability of reducing power for different conditions. The model does not resolve these mechanisms, so it cannot reproduce this pattern. This is an area of future research for the model.

The model was designed to be consistent with observed patterns, and the model-data comparison summarized above confirms that it behaves as designed (94% of the time). This comparison is important, because it is based on emerging behaviours of the model, not imposed rules (Grimm *et al.*, 2005; Hellweger, 2010; Koleva and Hellweger, 2015). For example, the heterocyst fraction is not imposed using a function that relates the number of heterocysts to the nutrient type, but it emerges from the complex interactions of the intracellular N assimilation system.

Using the model to interpret and analyze observations

The model helps interpret observations from experiments. For example, the nitrogenase level increases after about 24 h when cells are grown on fixed N and transferred to fixed N-free medium. Two papers show this pattern (Rowell *et al.*, 1977; Orr and Haselkorn, 1982) (Fig. 2E), and both suggest a slight temporary overexpression (overshoot) of nitrogenase. Neither of the authors of the original papers discuss this pattern, possibly because, in isolation, it is not clear if the pattern is due to experimental variability. However, the consistency among observations and the agreement with the model suggests that the pattern may reflect a real mechanism. In the model, this overshoot is a result of the time-delay between initiation of heterocyst differentiation and N fixation. Even while sufficient heterocysts are under development, the cells remain N stressed and continue to initiate heterocyst differentiation above the level required for these conditions. Other patterns that the model helped interpret include an initial decrease in N quota when light is turned on and a delay of NO3 uptake by NH4 and N2 grown cells (see Supporting Information).

The model can be used to quantitatively analyze observations and provide insights into the underlying mechanisms. For example, the model provides evidence for the existence of an N storage pool. Specifically, experiments with N2 isotopes show that the isotope level is initially higher in heterocysts and then higher in vegetative cells (Wolk *et al.*, 1974; Popa *et al.*, 2007) (Figs. 3A&B). This pattern can be explained qualitatively. Early on, the isotope level in heterocysts is higher because this is where N is fixed. Later, the pattern reverses because vegetative cells grow and assimilate the isotope into their biomass, whereas heterocysts do not grow. An earlier version of the model did not include a labile N pool (see LN in Fig. 1). When that model was applied to these experiments, it underestimated heterocyst isotope levels for all experiments (Fig. 3A&B, red lines). That is because the isotope in heterocysts was limited to the GLN and GLU pools, which are relatively small, consistent with observations [e.g. (Rowell *et al.*, 1977), see Supporting Information]. The model application showed that the higher isotope levels in heterocysts cannot be due to storage in GLN and GLU pools, which pointed to the existence of a storage reservoir. By including the LN pool, the model can reproduce the observed patterns (Fig. 3A&B, blue lines). Recent observations of N storage in cyanophycin granules in *Anabaena* heterocysts (Burnat *et al.*, 2014) are consistent with this finding. Other models also suggest a role for storage in timing of heterocyst development (Brown and Rutenberg, 2014). Figure 3C shows N fluxes in heterocysts and adjacent vegetative cells. In heterocysts, the net flux into the LN pool is negligible, but there is substantial exchange with the GLN and GLU pools.

A practical application of the model

Finally, we use the model to explore an ecologically-relevant question. The efficacy of N reduction to control lake eutrophication has been the subject of considerable debate (Schindler *et al.*, 2008; Paerl *et al.*, 2011). One argument is that reducing N input will not reduce primary productivity because cyanobacteria can simply turn on N fixation, which compensates for the reduction. The other argument is that N reduction will decrease productivity, because N fixation is a metabolically expensive process and is controlled by multiple environmental factors (besides P availability), including light and iron limitation, turbulence and ambient supersaturated dissolved oxygen concentrations that commonly occur in dense surface blooms (Moisander *et al.*, 2002; Paerl, 2009). Our model constitutes a unique opportunity to explore this question because it includes many of the mechanisms underlying growth on fixed and N2 gas. We simulated a population of *Anabaena* cells in a hypothetical lake with PO4 and NH4 in the inflow, subject to a number of nutrient reduction

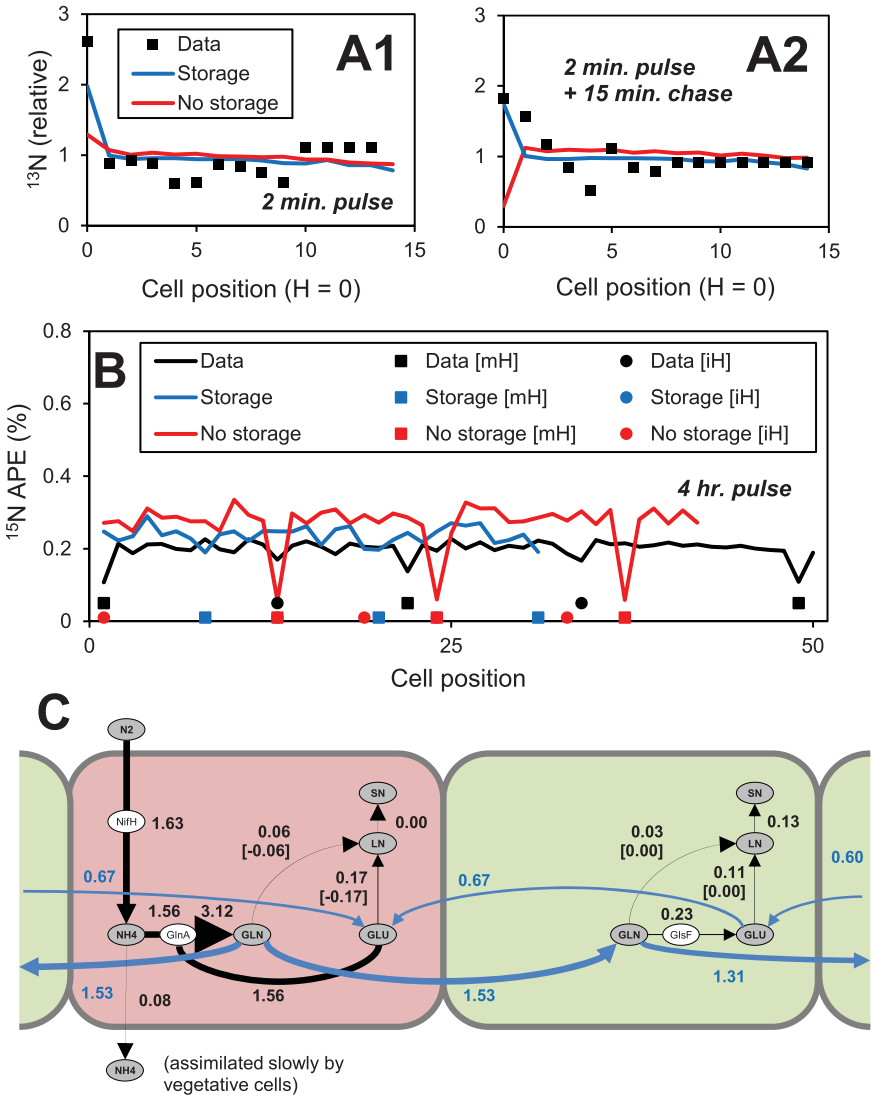

Fig. 3. Effect of N storage.
A&B. Distribution of N isotopes along the filament. Symbols are data from (Wolk *et al.*, 1974) and (Popa *et al.*, 2007). Lines are model results from full model (Storage) and model without labile N (LN) (No storage). In panel A, values are average ^{13}N levels vs. distance from heterocyst. In panel B, ^{15}N in a single filament is shown. Symbols mark the position of mature heterocyst (mH) and initiated heterocyst (iH). APE, atom percent enrichment.
C. N fluxes for growth on N2. Fluxes are in pmolN/cell/d. For fluxes between GLN/GLU and LN, the gross forward and backward (in '[]') fluxes are shown. See also Supporting Information Fig. S7.

scenarios, including 50% of P only, N only and N&P loading (Fig. 4). The P only reduction results in an approximately proportional decrease in chlorophyll *a*. For the N only scenario, the model predicts that N2 transcript and protein levels, heterocysts and N2 fixation increases following the reduction. However, the chlorophyll *a* concentration also decreases substantially, in this case by 33%. This scenario reflects the long-term field experiment conducted with Lake 227. Observations of chlorophyll *a* from this experiment are available, but the response due to the loading reduction is subject to debate (Scott and McCarthy, 2011). Previous phytoplankton models typically simulate N2 fixation in a simplified manner by eliminating N-limitation for species capable of N2 fixation (Wool *et al.*, 2013; Park and Clough, 2014), and they would not predict a decrease in biomass for this scenario. The largest effect is predicted for the dual P and N reduction; reducing N along with P results in an added 8% reduction. These predictions are a function of the lake properties (residence time, temperature, light and nutrient inflow concentrations), and the model does not account for other possible limiting factors that may occur in the field, such as turbulence and vertical mixing or the availability of trace nutrients (e.g. iron) and ambient dissolved oxygen concentrations, or the presence of non-fixing phytoplankton (Paerl, 2009). Therefore, this result should not be taken to apply uniformly across all lakes. However, our model suggests that substantial reduction in phytoplankton biomass can be obtained by reducing N alone or along with P.

Summary and outlook

We developed a dynamic, mechanistic, molecular-level model of *Anabaena* focusing on N interaction. The model

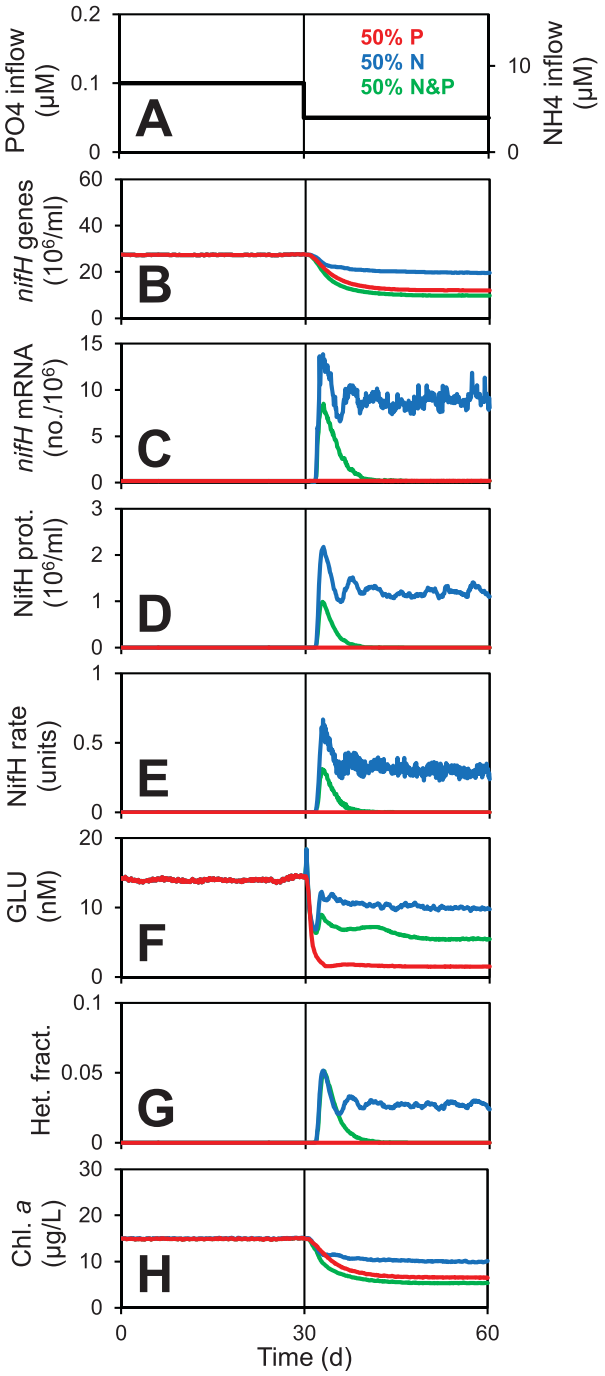

Fig. 4. Simulation of nutrient reduction scenarios in a hypothetical lake.
A. Inflow nutrient concentrations (i.e. loads).
B. *nifH* gene copies.
C. *nifH* transcript level (units are *nifH* transcripts per million total transcripts).
D. NifH protein copies.
E. NifH rate (μmolN2/l/d).
F. Glutamate.
G. Heterocyst fraction.
H. Chlorophyll *a* concentration.

was based on our current (although still simplified) understanding of the underlying processes, and informed and constrained by a database of observations from 55 different papers. Omitting inconsistencies in the observations, the model reproduces 94% of the 331 emerging patterns identified within these observations. The model helped to interpret some of the observations and was used to explore a practical question related to lake nutrient management.

There are a number of opportunities to improve the present model. It would be useful to validate the model by applying it, without any changes to the structure or parameters, to an independent database of experiments (we provide such a dataset in Supporting Information Section S2.65). The present model increases mechanistic detail and realism compared to previous models. However, it is still simple considering the present knowledge of *Anabaena*, and future work may increase the details of some features. For example, the N storage mechanism could be represented in more detail by breaking up or 'de-lumping' (Castellanos *et al.*, 2004; Hellweger, 2013) the existing labile N (LN) pool into cyanophycin, β-aspartyl-arginine, aspartate and arginine, and adding corresponding enzymes and fluxes, like the *all3922*-encoded isoaspartyl dipeptidase (Burnat *et al.*, 2014). We only model a small subset of genes, but the model can be extended to the genome scale, for which observations are available (Flaherty *et al.*, 2011; Park *et al.*, 2013). This extension could be done by adopting existing metabolic flux balance models (Nogales *et al.*, 2012) or proceeding in a more step-wise manner by 'de-lumping' several components (e.g. the nitrogenase complex). The database of observations also can be expanded. In this study, we restricted the model-data comparison to wild-type cells under natural conditions (i.e. those that may be encountered by the cells in the natural environment). However, laboratory experiments with inhibitors and mutant strains have been instrumental in increasing understanding of the mechanisms underlying N assimilation (Risser and Callahan, 2009; Burnat *et al.*, 2014). Those experiments can be used to further constrain the model (Hellweger, 2010, 2013).

A natural next step is to simulate a field population of *Anabaena* in a (real) lake. Field-scale *Anabaena* and agent-based phytoplankton models presented previously (Lewis *et al.*, 2004; Hellweger *et al.*, 2008; Bucci *et al.*, 2012) can be merged with the present model. The resulting model could predict dynamic gene, transcript, protein and metabolite levels (see Fig. 4) at the population and single-cell level, which can be compared with observations (Musat *et al.*, 2008; Oh *et al.*, 2011; Vila-Costa *et al.*, 2013; Hanson *et al.*, 2014) in the same manner as for the laboratory experiments (Fig. 2). The ability of the model to reproduce the observed behaviour of *Anabaena* in the

laboratory suggests the model cells will behave like real *Anabaena* cells in the field. Then, the model can be expanded to other species and processes. Application to other filamentous, N2-fixing species, like *Aphanizomenon* and *Nodularia*, would be straightforward and may only require recalibration of the parameters. The model can also be applied to non-N2-fixing species, like *Microcystis*, by knocking out *nifH* and related genes and recalibrating the parameters. Simulating toxin production would require adding additional genes and reactions, but the process is linked to N metabolism (Davis *et al.*, 2015), so the existing model already includes many related pieces (e.g. NtcA regulation). It remains to be determined if such a model will be feasible to set up for a real system and if it will provide better predictions than a simpler model. However, if that is the case, it will contribute significantly to research and management of lakes, such as Erie and Taihu.

This model summarizes the *Anabaena* – N interaction literature. As the amount of information on this subject grows, the format and size of a traditional review paper cannot capture the whole body of knowledge. Using a model constitutes a viable strategy to continue to summarize information at the system level. Advantages to this approach are summarizing observations in a quantitative manner and to aid in the translation of the knowledge. For example, managers can make use of this information by integrating this model into whole-lake ecosystem models. The (cyber-)infrastructure to support sharing such a model is increasingly available. This science approach is already being applied for other complex systems, such as in climate science, where models frequently are at the centre the discussion, driving data collection efforts to improve the model and using the model to make predictions.

Acknowledgements

We are grateful for funding from the National Science Foundation (NSF) (grant numbers H.W.P. NSF 1240851, S.W.W. NSF 1240870, W.S.G. and M.M. NSF 1240798, F.L.H. NSF 1240894). Thanks to Roger Giese (NU) for help interpreting enzyme assay data and Wolfgang Hess (U. Tuebingen) for interpreting his data. The development of this model benefited from interesting discussions with John Berges, Erica Young and Todd Miller (UWM). Four anonymous reviewers provided constructive criticism.

References

Ankrah, N.Y.D., May, A.L., Middleton, J.L., Jones, D.R., Hadden, M.K., Gooding, J.R., et al. (2014) Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *ISME J* **8:** 1089–1100.

Brown, A.I., and Rutenberg, A.D. (2014) A storage-based model of heterocyst commitment and patterning in cyanobacteria. *Phys Biol* **11:** 016001.

Bucci, V., Nunez-Milland, D., Twining, B., and Hellweger, F. (2012) Microscale patchiness leads to large and important intraspecific internal nutrient heterogeneity in phytoplankton. *Aquat Ecol* **46:** 101–118.

Burnat, M., Herrero, A., and Flores, E. (2014) Compartmentalized cyanophycin metabolism in the diazotrophic filaments of a heterocyst-forming cyanobacterium. *Proc Natl Acad Sci* **111:** 3823–3828.

Castellanos, M., Wilson, D.B., and Shuler, M.L. (2004) A modular minimal cell model: purine and pyrimidine transport and metabolism. *Proc Natl Acad Sci USA* **101:** 6681–6686.

Chapra, S.C. (1997) *Surface Water-Quality Modeling*. Boston: McGraw-Hill.

Davis, T.W., Bullerjahn, G.S., Tuttle, T., McKay, R.M., and Watson, S.B. (2015) Effects of increasing nitrogen and phosphorus concentrations on phytoplankton community growth and toxicity during planktothrix blooms in Sandusky Bay, Lake Erie. *Environ Sci Technol* **49:** 7197–7207.

Ehira, S., and Ohmori, M. (2006) NrrA, a nitrogen-responsive response regulator facilitates heterocyst development in the cyanobacterium Anabaena sp. strain PCC 7120. *Mol Microbiol* **59:** 1692–1703.

Ehira, S., Ohmori, M., and Sato, N. (2003) Genome-wide expression analysis of the responses to nitrogen deprivation in the heterocyst-forming cyanobacterium Anabaena sp. Strain PCC 7120. *DNA Res* **10:** 97–113.

Flaherty, B.L., Van Nieuwerburgh, F., Head, S.R., and Golden, J.W. (2011) Directional RNA deep sequencing sheds new light on the transcriptional response of Anabaena sp. strain PCC 7120 to combined-nitrogen deprivation. *BMC Genomics* **12:** 332.

Flores, E., and Herrero, A. (1994) Assimilatory nitrogen metabolism and its regulation. In *The Molecular Biology of Cyanobacteria*. Bryant, D., (ed). The Netherlands: Springer, pp. 487–517.

Flores, E., and Herrero, A. (2010) Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nat Rev Microbiol* **8:** 39–50.

Flynn, K.J., Fasham, M.J.R., and Hipkin, C.R. (1997) Modelling the interactions between ammonium and nitrate uptake in marine phytoplankton. *Philos Trans R Soc Lond Ser B: Biol Sci* **352:** 1625–1645.

Fuhrman, J., Follows, M., and Forde, S. (2013) Applying "-omics" Data in Marine Microbial Oceanography. *Eos, Trans Am Geophys Union* **94:** 241–241.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W.M., Railsback, S.F. et al. (2005) Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science* **310:** 987–991.

Hanson, B., Hewson, I., and Madsen, E. (2014) Metaproteomic survey of six aquatic habitats: discovering the identities of microbial populations active in biogeochemical cycling. *Microb Ecol* **67:** 520–539.

Hellweger, F.L. (2010) Resonating circadian clocks enhance fitness in cyanobacteria *in silico*. *Ecol Model* **221:** 1620–1629.

Hellweger, F.L. (2013) Escherichia coli adapts to tetracycline resistance plasmid (pBR322) by mutating endogenous potassium transport: in silico hypothesis testing. *FEMS Microbiol Ecol* **83:** 622–631.

Hellweger, F.L. (2015) 100 Years since Streeter and Phelps: it is time to update the biology in our water quality models. *Environ Sci Technol* **49:** 6372–6373.

Hellweger, F.L., Kravchuk, E.S., Novotny, V., and Gladyshev, M.I. (2008) Agent-based modeling of the complex life cycle of a cyanobacterium (Anabaena) in a shallow reservoir. *Limnol Oceanogr* **53:** 1227–1241.

Koleva, K.Z., and Hellweger, F.L. (2015) From protein damage to cell aging to population fitness in E. coli: insights from a multi-level agent-based model. *Ecol Model* **301:** 62–71.

Kreft, J.-U., Plugge, C.M., Grimm, V., Prats, C., Leveau, J.H.J., Banitz, T., et al. (2013) Mighty small: observing and modeling individual microbes becomes big science. *Proc Natl Acad Sci* **110:** 18027–18028.

Kumar, K., Mella-Herrera, R.A., and Golden, J.W. (2010) Cyanobacterial heterocysts. *Cold Spring Harbor Perspectives in Biology* **2:** a000315.

Layzell, D.B., Turpin, D.H., and Elrifi, I.R. (1985) Effect of N source on the steady state growth and N assimilation of P-limited Anabaena flos-aquae. *Plant Physiol* **78:** 739–745.

Lee, D.-Y., and Rhee, G.Y. (1997) Kinetics of cell death in the cyanobacterium anabaena FLOS-aquae and the production of dissolved organic carbon1. *J Phycol* **33:** 991–998.

Lee, D.-Y., and Rhee, G.Y. (1999) Kinetics of growth and death in anabaena FLOS-aquae (cyanobacteria) under light limitation and supersaturation. *J Phycol* **35:** 700–709.

Lewis, D., Brookes, J., and Lambert, M. (2004) Numerical models for management of Anabaena circinalis. *J Appl Phycol* **16:** 457–468.

Meeks, J.C., Wolk, C.P., Thomas, J., Lockau, W., Shaffer, P.W., Austin, S.M., et al. (1977) The pathways of assimilation of 13NH4+ by the cyanobacterium, Anabaena cylindrica. *J Biol Chem* **252:** 7894–7900.

Moisander, P.H., Hench, J.L., Kononen, K., and Paerl, H.W. (2002) Small-scale shear effects on heterocystous cyanobacteria. *Limnol Oceanogr* **47:** 108–119.

Muro-Pastor, A.M., and Hess, W.R. (2012) Heterocyst differentiation: from single mutants to global approaches. *Trends Microbiol* **20:** 548–557.

Musat, N., Halm, H., Winterholler, B., Hoppe, P., Peduzzi, S., Hillion, F., et al. (2008) A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. *Proc Natl Acad Sci* **105:** 17861–17866.

Nogales, J., Gudmundsson, S., Knight, E.M., Palsson, B.O., and Thiele, I. (2012) Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proc Natl Acad Sci* **109:** 2678–2683.

Oh, S., Caro-Quintero, A., Tsementzi, D., DeLeon-Rodriguez, N., Luo, C., Poretsky, R., and Konstantinidis, K.T. (2011) Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. *Appl Environ Microbiol* **77:** 6000–6011.

Ohmori, M., Ohmori, K., and Strotmann, H. (1977) Inhibition of nitrate uptake by ammonia in a blue-green alga, Anabaena cylindrica. *Arch Microbiol* **114:** 225–229.

Orr, J., and Haselkorn, R. (1982) Regulation of glutamine synthetase activity and synthesis in free-living and symbiotic Anabaena spp. *J Bacteriol* **152:** 626–635.

Paerl, H. (2009) Controlling eutrophication along the freshwater–marine continuum: dual nutrient (N and P) reductions are essential. *Estuaries Coasts* **32:** 593–601.

Paerl, H.W. (1984) Transfer of N2 and CO2 fixation products from anabaena oscillarioides to associated bacteria during inorganic carbon sufficiency and deficiency1. *J Phycol* **20:** 600–608.

Paerl, H.W., Xu, H., McCarthy, M.J., Zhu, G., Qin, B., Li, Y., and Gardner, W.S. (2011) Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): the need for a dual nutrient (N & P) management strategy. *Water Res* **45:** 1973–1983.

Park, J.-J., Lechno-Yossef, S., Wolk, C.P., and Vieille, C. (2013) Cell-specific gene expression in Anabaena variabilis grown phototrophically, mixotrophically, and heterotrophically. *BMC Genomics* **14:** 759.

Park, R.A., and Clough, J.S. (2014) *Aquatox (Release 3.1 plus), Modeling Environmental Fate and Ecological Effects in Aquatic Ecosystems, Volume 2: Technical Documentation*. U.S. Environmental Protection Agency (USEPA), Washington, DC, p. 354.

Popa, R., Weber, P.K., Pett-Ridge, J., Finzi, J.A., Fallon, S.J., Hutcheon, I.D., et al. (2007) Carbon and nitrogen fixation and metabolite exchange in and between individual cells of Anabaena oscillarioides. *ISME J* **1:** 354–360.

Ramos, J.L., Guerrero, M.G., and Losada, M. (1984) Sustained photoproduction of ammonia from dinitrogen and water by the nitrogen-fixing cyanobacterium anabaena sp. strain ATCC 33047. *Appl Environ Microbiol* **48:** 114–118.

Rhee, G.Y., and Lederman, T.C. (1983) Effects of nitrogen sources on p-limited growth of anabaena FLOS-aquae1. *J Phycol* **19:** 179–185.

Risser, D.D., and Callahan, S.M. (2009) Genetic and cytological evidence that heterocyst patterning is regulated by inhibitor gradients that promote activator decay. *Proc Natl Acad Sci* **106:** 19884–19888.

Rowell, P., Enticott, S., and Stewart, W.D.P. (1977) Glutamine synthetase and nitrogenase activity in the blue-green alga anabaena cylindrica. *New Phytol* **79:** 41–54.

Schindler, D.W., Hecky, R.E., Findlay, D.L., Stainton, M.P., Parker, B.R., Paterson, M.J., et al. (2008) Eutrophication of lakes cannot be controlled by reducing nitrogen input: results of a 37-year whole-ecosystem experiment. *Proc Natl Acad Sci* **105:** 11254–11258.

Scott, T.J., and McCarthy, M.J. (2011) Response to comment: nitrogen fixation has not offset declines in the lake 227 nitrogen pool and shows that nitrogen control deserves consideration in aquatic ecosystems. *Limnol Oceanogr* **56:** 1548–1550.

Steffen, M.M., Belisle, B.S., Watson, S.B., Boyer, G.L., and Wilhelm, S.W. (2014a) Status, causes and controls of cyanobacterial blooms in Lake Erie. *J Great Lakes Res* **40:** 215–225.

Steffen, M.M., Dearth, S.P., Dill, B.D., Li, Z., Larsen, K.M., Campagna, S.R., and Wilhelm, S.W. (2014b) Nutrients drive transcriptional changes that maintain metabolic homeostasis but alter genome architecture in Microcystis. *ISME J* **8:** 2080–2092.

Subramanian, G., and Shanmugasundaram, S. (1986) Uninduced ammonia release by the nitrogen-fixing cyanobacterium Anabaena. *FEMS Microbiol Lett* **37:** 151–154.

Vila-Costa, M., Sharma, S., Moran, M.A., and Casamayor, E.O. (2013) Diel gene expression profiles of a phosphorus limited mountain lake using metatranscriptomics. *Environ Microbiol* **15:** 1190–1203.

Wolk, C.P., Austin, S.M., Bortins, J., and Galonsky, A. (1974) Autoradiographic localization of 13N after fixation of 13N-labeled nitrogen gas by a heterocyst-forming blue-green alga. *J Cell Biol* **61:** 440–453.

Wool, T.A., Robert B. Ambrose, J., and Martin, J.L. (2013) *WASP7 – Multi-Algal Model Theory and User's Guide*. Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development, p. 21.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Model technical documentation and model-data comparison

Movie S1. Simulation of a single filament growing in a turbidostat with changing N inflow (NH4 $>$ none).