Extreme bioengineering to meet the nitrogen challenge

Stefan Burén[a], Gema López-Torrejón[a,b], and Luis M. Rubio[a,b,1]

In the 1970s, investigators were already envisioning new technologies to generate crops that could utilize nitrogen (N_2) from air to produce their own fixed N species, stating that "cereals that could provide their own fertilizer are beyond doubt the biggest prize of all in the gift of the new biology" (1). At this time, the N_2-fixation genes (*nif*) of *Klebsiella pneumoniae* had just been transferred to *Escherichia coli*, making the naturally nonfixing bacterium capable of growing in the absence of combined N (2) and giving hope that this feature could also be carried over to more complex organisms. In PNAS, almost 50 years after the generation of the first N_2-fixing *E. coli*, researchers from the Peking University, in collaboration with the lead author from that breakthrough study, present an ingenious approach to achieve coordinated and stoichiometric expression of *nif* genes in heterologous hosts (3). The method is based on regrouping genes of complex biological pathways into synthetic "giant genes," products of which may be further processed into shorter functional proteins on demand. This technology will revolutionize our attempts to transfer the elaborate *nif* machinery into heterologous hosts, providing researchers with a groundbreaking advance to help generate plants capable of self-fertilization (4, 5).

Successful transfer of *nif* genes into nonfixing organisms, such as plants, will depend on our ability to execute rational simplifications. Researchers must aim at finding the least possible complex, and most possible robust, genetic components capable of performing nitrogenase-related reactions in cellular environments different from their natural hosts. We must also identify intermediate steps of nitrogenase biosynthesis and develop methods for their analysis so that nitrogenase can be stepwise engineered—and functionally verified—before integration into a complete pathway. Once these objectives are achieved, synthetic biology, combined with increased knowledge of nitrogenase biogenesis and catalysis, will enable fine-tuning and improvement of complete N_2-fixing reaction.

Reasons for the need of simplification are plentiful. Genes that produce soluble, stable, and functional proteins in eukaryotic target hosts must be identified. Expression and regulation of multiple *nif* genes must be controlled, as they are naturally tightly regulated (6, 7). Functionality of some Nif components depend on the correct formation of multimeric complexes, such as NifDK and NifEN; others form more transient interactions that assist in Nif protein folding, protection, and delivery of cofactors or electrons (8). Finally, nitrogenase functionality depends on the biophysical environment of the new cellular host. Well-known factors affecting nitrogenase include sensitivity of its metal clusters to oxygen exposure, ATP levels and reducing power required for the nitrogenase reaction, and metal availability and homeostasis for the biosynthesis of the various nitrogenase cofactors (9, 10).

Reducing nitrogenase genetic requirements to a minimum will likely be crucial for transfer to eukaryotic hosts. Current advances in synthetic biology have provided instruments to pursue the ideas put forward in the 1970s, and recent studies have shown how these new tools can rewrite and refactor naturally occurring *nif* gene clusters. In excellent studies involving *K. pneumoniae* or *Klebsiella oxytoca* as *nif* sources, gene clusters were reorganized or refactored and transferred to *E. coli*. Nonessential genetic parts could be cleaned up, and native *nif* regulation was replaced with fully defined expression units that enabled analysis and optimization of the heterologous *E. coli* nitrogenase (11–14).

In another study published in PNAS last year by these authors (15), synthetic biology was employed to reconstruct electron-transport components consisting of ferredoxin-NADPH oxidoreductases and ferredoxins originating from plant organelles. Their ability to feed electrons to nitrogenase was tested to identify plant or plant-related proteins that could perform some of the reactions needed for nitrogenase to reduce the number of prokaryotic genes needed to generate a putative N_2-fixing plant.
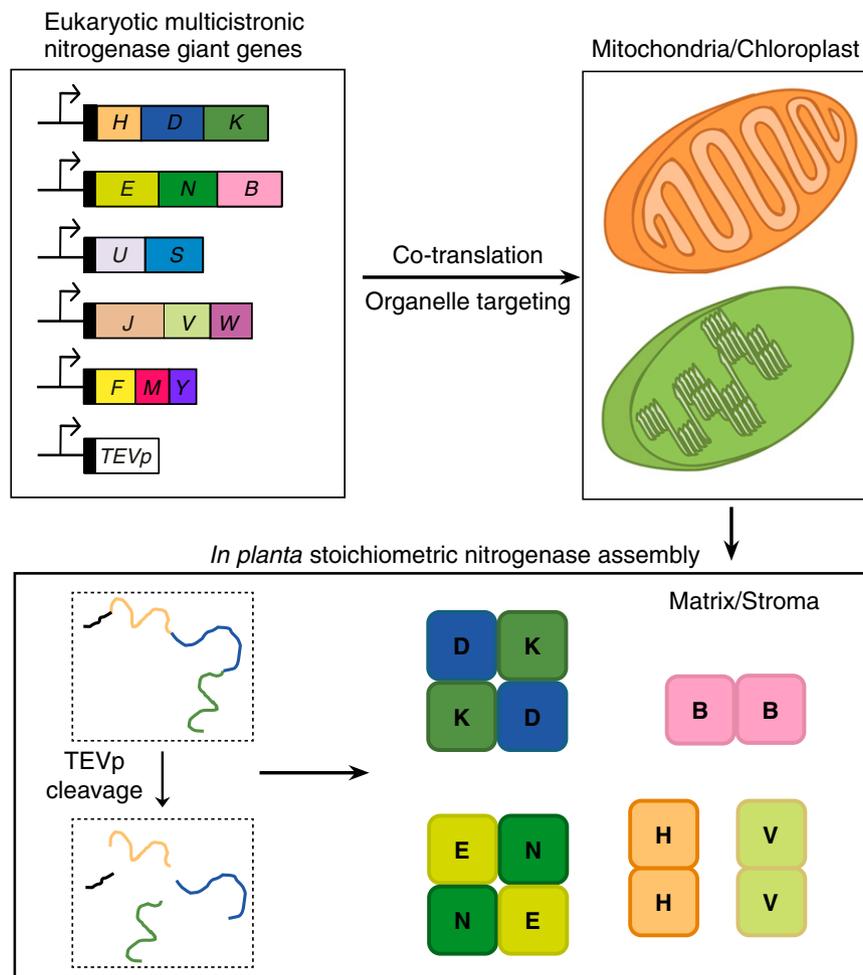
Fig. 1. Prospects for giant genes in plant nitrogenase engineering.

In recent years, we have also seen attempts to express nitrogenase components in eukaryotic cells (e.g., *Saccharomyces cerevisiae* and *Nicotiana benthamiana*), and in vitro activity of purified yeast NifH and NifB proteins were reported (16–19). NifH was only active upon targeting to the mitochondria when cultured cells were grown under aerobic conditions, strengthening the assumption that ATP production and O_2 consumption will be important for the functionality of plant nitrogenase (20). Unfortunately, while mitochondrial targeting minimizes O_2 exposure, it introduces another challenge, as the proteins now have to be equipped with a leader sequence that must be properly cleaved upon mitochondrial import. In the perhaps most elaborate eukaryotic Nif study published so far, nine *nif* genes (*nifHDKUSMENB*) were cotransferred to the genome of yeast (21). In contrast to the other two studies that used strong and inducible galactose-dependent promoters to drive Nif expression, constitutive promoters were designed here to render expression levels and stoichiometries more similar to the natural *nif* system (6). From this work, it became clear that accumulation of Nif proteins in mitochondria did not always correlate to expected expression levels and that mitochondria leader sequences could influence Nif protein accumulation. Importantly, expression of NifDK and NifEN heterotetramers proved difficult, indicating that complex formation upon import to the mitochondria was challenging.

To reduce the number of genes transferred and to enable stoichiometric expression of Nif components in eukaryotic cells, internal ribosome entry site (IRES) technology or so-called 2A

peptides can be employed to produce multiple proteins from single transcripts. While IRES initiates translation in a cap-independent manner, the 2A peptide impairs formation of the peptide bond. As translation continues, it results in "ribosomal skipping" or "cleavage." The 2A peptide provides expression at equimolar levels, but it also leaves an approximately 20-amino acid extension at the carboxyl terminus of the protein upstream of the 2A peptide and an extra proline residue at the amino terminus of the protein downstream. These extensions can impair functionality or stability of proteins. While 2A peptides and IRES have successfully been used in mammalian cells—for example, for the creation of multicistronic gene-expression cassettes capable of generating induced pluripotent stem cells (22)—targeting and cleavage efficiency in plants have been questioned (23).

In the PNAS article, Yang et al. (3) provide an ingenious methodology that could overcome many of the above-listed hurdles. Instead of using IRES or 2A peptides, the authors exploit a post-translational protein splicing strategy based on the tobacco etch virus protease (TEVp). TEVp specifically cleaves polypeptides harboring the ENLYFQS sequence (at Q\S), leaving shorter overhangs (6 and 1 amino acids at the carboxyl and amino termini, respectively) than 2A peptides and enabling multiple genes to be fused and expressed as giant polypeptides that subsequently generate individual proteins by the action of TEVp. First, Yang et al. (3) analyzed expression levels of 18 *nif* genes endowing *E. coli* with N_2 fixation to assign them to subgroups. Then, the "tailing tolerance"

was determined for each Nif protein. Taking into consideration the native expression levels and tailing tolerance and further eliminating four *nif* genes that proved to be nonessential for *E. coli* (*nifTXZQ*), five giant genes were constructed (*nifHĎDĎK*, *nifEĎN~B*, *nifUĎS*, *nifJĎVĎW*, and *nifFĎMĎY*). Expression of these giant genes in *E. coli* resulted in 72% of nitrogenase activity relative to the original system and supported diazotrophic growth of the bacterium. Quantitative analysis of individual protein levels revealed that stoichiometry of most components from this polyprotein-based expression system correlated well with the reconstituted operon-based system in *E. coli* and with the native system in the original *K. oxytoca* host.

The Yang et al. work (3) provides a conceptual advance to generate N_2-fixing plants by describing a method for (Nif) protein synthesis from multicistronic genes with much shorter carboxyl-terminal overhang than would be the case from using 2A peptides (Fig. 1). While this reduces the tailing effect, even more important is that it facilitates organelle targeting, as would be the case if nitrogenase were to be located at the plant mitochondrial matrix. Because both IRES and 2A peptide results in "cleavage" at the site of translation (cytoplasm), subsequent mitochondria/chloroplast targeting would require targeting signals for each and every cleaved polypeptide. However, using TEVp, only one leader sequence will be needed per giant gene, facilitating cloning in plant expression vectors and transformation of plant genomes. In addition, because cleavage would occur within the organelle itself (where TEVp is targeted), folding of multimeric Nif proteins (e.g., NifDK and NifEN) will likely be improved. This technology could also help taking advantage of previous bacterial *nif* refactoring work (11–14) by creating eukaryotic "operons" that minimize the number of required transcriptional units, thus making *nif* gene expression more robust and easier to optimize. It will be crucial to test how well it performs in eukaryotic cells and importantly upon organelle targeting. TEVp must not be active outside the organelle because it would prevent import of the proteins downstream of the primary TEVp recognition site. If successful in eukaryotic organelles, this technology will be a milestone in the path to N_2-fixing plants.

1 Hardy RW, Havelka UD (1975) Nitrogen fixation research: a key to world food? *Science* 188:633–643.
2 Dixon RA, Postgate JR (1972) Genetic transfer of nitrogen fixation from *Klebsiella pneumoniae* to *Escherichia coli*. *Nature* 237:102–103.
3 Yang J, et al. (2018) Polyprotein strategy for stoichiometric assembly of nitrogen fixation components for synthetic biology. *Proc Natl Acad Sci USA* 115:E8509–E8517.
4 Good A (2018) Toward nitrogen-fixing plants. *Science* 359:869–870.
5 Burén S, Rubio LM (2018) State of the art in eukaryotic nitrogenase engineering. *FEMS Microbiol Lett* 365:fnx274.
6 Poza-Carrión C, Jiménez-Vicente E, Navarro-Rodríguez M, Echavarri-Erasun C, Rubio LM (2014) Kinetics of Nif gene expression in a nitrogen-fixing bacterium. *J Bacteriol* 196:595–603.
7 Dixon R, Kahn D (2004) Genetic regulation of biological nitrogen fixation. *Nat Rev Microbiol* 2:621–631.
8 Jimenez-Vicente E, et al. (2018) Sequential and differential interaction of assembly factors during nitrogenase MoFe protein maturation. *J Biol Chem* 293:9812–9823.
9 Rubio LM, Ludden PW (2008) Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Annu Rev Microbiol* 62:93–111.
10 Hoffman BM, Lukoyanov D, Yang Z-YY, Dean DR, Seefeldt LC (2014) Mechanism of nitrogen fixation by nitrogenase: The next stage. *Chem Rev* 114:4041–4062.
11 Wang X, et al. (2013) Using synthetic biology to distinguish and overcome regulatory and functional barriers related to nitrogen fixation. *PLoS One* 8:e68677.
12 Temme K, Zhao D, Voigt CA (2012) Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proc Natl Acad Sci USA* 109:7085–7090.
13 Smanski MJ, et al. (2014) Functional optimization of gene clusters by combinatorial design and assembly. *Nat Biotechnol* 32:1241–1249.
14 Yang J, Xie X, Wang X, Dixon R, Wang Y-P (2014) Reconstruction and minimal gene requirements for the alternative iron-only nitrogenase in *Escherichia coli*. *Proc Natl Acad Sci USA* 111:E3718–E3725.
15 Yang J, Xie X, Yang M, Dixon R, Wang Y-P (2017) Modular electron-transport chains from eukaryotic organelles function to support nitrogenase activity. *Proc Natl Acad Sci USA* 114:E2460–E2465.
16 López-Torrejón G, et al. (2016) Expression of a functional oxygen-labile nitrogenase component in the mitochondrial matrix of aerobically grown yeast. *Nat Commun* 7:11426.
17 Burén S, Jiang X, López-Torrejón G, Echavarri-Erasun C, Rubio LM (2017) Purification and in vitro activity of mitochondria targeted nitrogenase cofactor maturase NifB. *Front Plant Sci* 8:1567.
18 Allen RS, et al. (2017) Expression of 16 nitrogenase proteins within the plant mitochondrial matrix. *Front Plant Sci* 8:287.
19 Pérez-González A, et al. (2017) Adaptation of the GoldenBraid modular cloning system and creation of a toolkit for the expression of heterologous proteins in yeast mitochondria. *BMC Biotechnol* 17:80.
20 Curatti L, Rubio LM (2014) Challenges to develop nitrogen-fixing cereals by direct *nif*-gene transfer. *Plant Sci* 225:130–137.
21 Burén S, et al. (2017) Formation of nitrogenase NifDK tetramers in the mitochondria of *Saccharomyces cerevisiae*. *ACS Synth Biol* 6:1043–1055.
22 Sommer CA, et al. (2009) Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells* 27:543–549.
23 Samalova M, Fricker M, Moore I (2006) Ratiometric fluorescence-imaging assays of plant membrane traffic using polyproteins. *Traffic* 7:1701–1723.

Burén et al.