Supplement to Xu, Elhai, and Wolk (2007)
Methods used to analyze data of Ehira and Ohmori (2006)

Apart from the methods described in the article itself the following may illuminate how the results presented in Tables 14.1 through 4 and Figures 14.2 and 3 were obtained.

In the experiments of Ehira and Lechno-Yossef and their coworkers (Ehira and Ohmori, 2006; Ehira et al., 2003; Lechno-Yossef et al., 2006), constant amounts of RNA were loaded per microarray. Because a very large fraction of the mass of the RNA in a cell is stable RNA (rRNA and tRNA; Dennis and Bremer, 1974), this procedure effectively, albeit crudely, normalized RNA levels detected by their microarrays to stable RNA. Unfortunately, in many organisms, nitrogen deprivation triggers cessation of rRNA synthesis (Dennis et al., 2004) and considerable degradation of stable RNA (Kaplan and Apirion, 1975). Therefore, in the absence of measurements of RNA levels per cell or per volume of culture, expression ratios determined by the microarrays under consideration are suspect. An apparent two-fold increase of expression following nitrogen deprivation might be the result of the loss of 50% of the culture's stable RNA. Without proper normalization it is not possible even to say with assurance whether the abundance of a transcript went up or down, let alone by what degree.

Normalizing individual signals to total signals resulting from mRNA would solve the problem if the overall abundance of mRNA did not change (Rhodius et al., 2002), but we know of no reason to think that this is a valid procedure in the case of nitrogen deprivation. The assumption that genes whose transcription changes constitute a minor part of the whole may be tenable in many experiments, but one can imagine that nitrogen deprivation might provoke an overall drop in transcription in an attempt to conserve nitrogenous metabolites or an overall increase, owing to a large number of genes that are induced and are related to heterocyst differentiation (Wealand et al., 1989). Another popular procedure is to normalize signals to transcripts of one or more "housekeeping genes" presumed to remain stable over the course of the experiment. However, stable expression has been demonstrated for no gene during heterocyst differentiation. One commonly used control gene, *rnpB* (Vioque, 1992), which encodes the RNA component of RNaseP, may be a reasonable surrogate measure of total RNA, but as discussed above, total RNA may be a poor normalization standard. In *E. coli*, *rnpB* has been shown to be regulated by amino acid deprivation (Jung and Lee, 1997), precisely the behavior we would like to avoid in a standard.

Since we have no faith that levels of either total RNA or total mRNA remain constant when *Anabaena* is shifted to nitrogen-depleted conditions, we sought an assumption that seemed safer. Our assumption is that most genes will not increase or decrease expression per cell unless there is a global regulatory mechanism or condition that affects all or nearly all genes. Putting that possibility aside, then if fewer than 50% of genes have reduced expression per cell and fewer than 50% of genes have increased expression per cell after a shift to –N conditions, then the median expression ratio after the shift will be 1. We have relied on that assumption in normalizing the data. Note that if this assumption is invalid, then surely the assumption used by Ehira and Ohmori (2006) is as well.

We normalized the data of Ehira and Ohmori (2006) in two stages:

1. In accord with Ehira and Ohmori, we normalized each channel by the total fluorescent intensity of all spots, after subtracting background intensity for each spot. At this stage, the partially normalized intensity for a given spot is:

$$\frac{100 * (\text{spot intensity} - \text{spot background intensity})}{\text{average of (spot intensity} - \text{spot background intensity)}}$$

   Note that since the microarrays of Ehira and Ohmori did not include stable RNA (e.g. rRNA), the average is effectively of mRNA.

2. The intensity was further adjusted by dividing the partially normalized intensity by the median expression ratio (partially normalized intensity after the shift divided by the partially normalized intensity before the shift). So the fully normalized intensity is given as:

   partially normalized intensity / median intensity ratio of genes at time *t*

   The median intensity ratio was calculated only amongst those genes that had significant expression under all three experimental conditions, as defined by Ehira and Ohmori (see below). There were 1539 such genes. By definition, the partially normalized intensities of gene expression under +N conditions were left unchanged.

The result is that the fully normalized average gene expression of a gene under +N conditions is set at 100. The fully normalized average gene expression of a gene under other conditions may be more or less than this. The median gene expression ratio is 1 under all experimental conditions.

We adopted the definition of "significantly above background" used by Ehira and Ohmori (2006). Genes were considered above background if and only if the fluorescence intensity of no more than one replicate fell below 1.5 times the fluorescence intensity of the background for that spot. This procedure removed from consideration 34% to 60% of the genes examined, depending on the experimental condition. We also excluded on of the six replicates of the experiment where nitrogen was deprived for 8 hours (Ehira and Ohmori's first replicate), because values for many genes are missing. The other 17 replicates have values for the same 5336 genes of *Anabaena*.