**Aim 3: Metabolic Regulation of Asn: Correlating the Transcriptome with the Metabolome**

**Rationale:** We aim to integrate our studies on transcriptional control of Asn synthesis/metabolism into a regulatory network model that includes N-assimilation metabolites. In Aim 3A we ask: Can changes in TF expression affect levels of Asn and related metabolites? For example, RAV2 is predicted to induce Asn synthesis and also block Asn catabolism (Fig. X). We will test whether 35S::RAV2 transgenics have a greater effect on improving seed N content, compared to our previous success with 35S::ASN1 transgenics (Lam et al., 2003). In Aim 3B we ask: Do changes in N-flux into Asn affect N-regulated gene transcription? Signaling roles for both Gln and Asn are supported by our studies in Arabidopsis, see Progress and (Gutierrez et al 2008, Lam et al 1998). The Gln/Asn ratio was also suggested to influence regulation of genes involved in the synthesis of the major amino acids (Gln, Glu, Asp and Asn) in corn cobs (Seebauer et al 2004). We will use ASN1 and ANS1 mutants/transgenics to test whether *in vivo* changes in N-flux into Asn affects organic N response. Following various N-treatments (NO3-, Gln, Asn) of ASN1/ANS1 mutants/transgenics vs. wild-type, we will perform a combined amino acid profiling and transcriptional analyses. We will use *machine learning by stochastic gradient descent with regularization* to infer relationships between transcriptional control and metabolic output of the N-assimilation pathway. This analysis will globally test the hypothesis that the Gln/Asn ratio alters the regulation of gene expression, and also derive new hypotheses for genome-wide metabolic regulation in response to such N-metabolite signaling.

**Aim 3A: Transcriptional Regulation of Asn Metabolism**. The goal of this aim is to determine whether transcription factors (TFs) that target ASN1/ANS1 affect changes in N-flux into Asn. This will have practical application for altering N-assimilation in plants. Supplying stable isotope labeled 15NO3- will allow us to follow newly assimilated nitrogen. As Asn is preferentially synthesized under conditions of N-excess and C-limitation (dark), we will optimize the treatment conditions (Progress Aim 3). Following pre-treatment in ammonium succinate, we will increase 15N-treatment concentration to 5mM 15NO3- and harvest plants at extended time intervals (2, 6, 12, 18 and 24 h) following transfer to light vs. dark. Using conditions that optimize Asn accumulation, we will track15N-assimilation into (Asp, Gln, Asn, Glu) the substrates/products of Asn synthesis/metabolism. Mutants/35S::transgenics of TFs predicted to reciprocally affect ASN1 and ANS1 (ANAC047, ANAC102, and RAV2) (see Aim 2B) will be tested. We currently have homozygous T-DNA insertion SALK lines for ANAC047 (SALK\_066615), ANAC102 (SALK\_030702), and RAV2 (SALK\_070847), as well as seeds of 35S::RAV2 and 35S::ANAC102 overexpressor lines (Castillejo and Pelaz, 2008; Christianson et al., 2009). We are in the process of generating a 35S::ANAC047 overexpressor line using a Gateway cloning strategy. Mutants will be grown, treated, and harvested for metabolite extraction and GC-MS analysis (as described in Progress Aim 3) to determine the effect of these TFs on 15N-labeled amino acid metabolite levels in seedlings. Mutants/transgenics that show alterations in Asn levels in seedlings, will be analyzed for total seed-N, in collaboration with Dr. HM Lam (Lam et al., 2003).

**Expected Results and Potential Pitfalls:** This aim will provide us with information on whether and how the TFs RAV2, ANAC047 and ANAC102 influence levels of Asn and seed N, which would have practical implications. Due to genetic redundancy, we may not see significant phenotypic changes in single T-DNA mutations, and we are more likely to have greater success with 35S::overexpressor lines. For example, we predict that 35S::RAV2 will increase Asn levels, both by inducing Asn synthesis (via ASN1) *and* by repressing Asn metabolism (via ANS1), which would be a putuative commercial target. It is noteworthy that high yielding corn lines have a QTL that is associated with reduced expression of asparaginase in corn (Moose et al. XXXX) (NSF Plant Genome, PI meeting abstract, 2010).

**Aim 3B: Metabolic Regulation of Transcription**

**Rationale**: In this subaim, we will explore the role of Asn as a potential feedback “signal” in the regulation of Asn synthesis and metabolism, a hypothesis for which we have preliminary data as described in Aim 3B of the Progress Report. We aim to experimentally show how alterations in the *in vivo* levels of Asn (using T-DNA and 35S::overexpressor lines for ASN1 and ANS1) affect whole plant N responses, and to better define the role of Asn in N-signaling (Figure X). By performing N-treatments on plants with alterations in ASN1/ANS1, we will determine how and whether perturbations in N-assimilation into endogenous Asn *in vivo*, affects gene regulation. Using a method called “gradient descent” to infer relationships between transcriptional control and metabolic output of the N-assimilation pathway, we will globally test the hypothesis that the Gln/Asn ratio alters the regulation of gene expression, and also derive new hypotheses for genome-wide metabolic regulation in response to such N-assimilate signaling

**Mutants/Transgenics in Asparagine synthetase (ASN1) and asparaginase (ANS1) for metabolite analysis.** ASN1 and ANS1 are the primary isoforms involved in the synthesis and degradation of free and transported Asn, as shown by transgenic and enzymatic studies (Lam et al., 2003; Bruneau et al., 2006). We will use T-DNA insertion and 35S-overexpressor lines of ASN1 and ANS1 to alter endogenous levels of Asn and determine effects on transcription compared to WT. The endogenous levels of Asn in WT, mutant, and transgenic plants will be determined by GC-MS analysis as described in Aim 3 of the Progress Report. For these studies, we will look only at total unlabeled 14N pools. We have a homozygous gene-trap line for ASN1 (GT9012) from the [Martienssen lab](http://www.cshl.org/public/overviews/martienssen.html) at Cold Spring Harbor, and isolated a homozygous T-DNA insertion line for ANS1 (SALK\_074531). We previously generated a 35S::ASN1 line (Lam et al., 2003) and are currently generating a 35S::ANS1 overexpressor line using Gateway cloning technology. We previously showed that Asn levels are elevated in the 35S::ASN1 line (Lam et al., 2003) and we are currently processing ASN1 gene-trap mutant samples for metabolite analysis to determine if Asn levels are decreased. Transgenic/mutant lines of ASN1 and ANS1 that show changes in levels of endogenous Asn will be used in the sections below, to study how altering N-assimilation into endogenous Asn affects N-signaling *in vivo*.

***a) Collection of Transcriptomic and Metabolite data:*** We will simultaneously collect metabolite and transcriptional data from wild-type, mutant and 35S-transgenic ASN1 and ANS1 plants (described above) for co-analysis to identify significant transcriptional changes in response to changes in endogenous Asn levels. Experimentally, growth, pre-treatment and treatment conditions will be the same as described in the Progress Report, except that plants will be treated with either 1 mM KNO3, 3.4 mM Gln, 0.4mM Asn or 1 mM KCl (as control) in light or dark conditions (Fig. X). These concentrations of Gln and Asn were previously determined to affect expression of the ASN1 gene, without having a detrimental effect on plant growth (Lam et al 1998). Collected shoot tissue will be separated for parallel analysis of transcripts and metabolites. Amino acids will be analyzed by GC-MS as described in Progress Aim 3A, and quantified as nmol/g FW. We will initially determine the expression levels of genes that have predicted associations with ASN1 and ANS1 by qPCR, as they are likely candidates for influence by Asn signaling (e.g. RAV2, see Progress Aim 3B). Microarray analysis will be used to correlate genome-wide transcriptional changes with endogenous levels of Gln vs Asn. This analysis will address whether and how altering the Gln/Asn ratio in source tissue affects gene expression. This Gln/Asn ratio has been proposed to be a N-signal to affect kernel development in corn (Seebauer et al., 2004). This analysis will also attempt to provide additional support for our hypothesis that Asn is a signal molecule by supplying ASN1 mutant plants with exogenous Asn.

***b) Analysis and Integration of Metabolic and Transcriptional Data:*** The quantitative metabolite and genome-wide transcriptional data from wild-type and mutants/transgenics will be analyzed by: i) ANOVA with model-reduction, and ii) machine learning by stochastic gradient descent with regularization. These methods of analysis will provide us with detailed quantitative information about metabolite and transcript interactions. In both cases, we will attempt to obtain an equation that determines the expression of each gene and level of amino acid. This linear equation will depend on NO3 levels, Gln, Asn levels, and interaction terms with the form of the Asn mutant (corresponding to Asn levels 0, 1, and 2 in Mutant, WT, and Overexpressor (OX) plants, respectively). Specifically, the variables are NO3 level, Gln level, Asn level, NO3\*Asn, Gln\*Asn (and others depending on the conditions tried). The statistical technique assumes an underlying normal distribution. The gradient descent technique assumes that a magnitude based regularization term gives a good outcome. We will test both outcomes using bootstrap cross-validation on the experiments.

In our ANOVA modeling, the response variable will be either the expression of a gene or the level (abundance) of a metabolite. The explanatory variables include 3 factors: N treatment, with 4 levels, KCl, KNO3, Gln, Asn; Light treatment, with 2 levels, Dark or Light; and Genotype, with 3 possible levels, WT, mutant and OX. For each gene/metabolite we will start with the most complex model, which occludes main effects, 2-way, and 3-way interactions. We will simplify each model in a step-wise, hierarchical fashion and check for statistical justification of each simplification (Crawley, 2007) until we have found the simplest model that retains explanatory power. Integer coding of factor levels allows the flexibility of simplifying a model by removing any desired interaction term if it is deemed to be not significant. We will consider the trio (WT, KCl, Dark) to be control combination of factor levels. The terms in the final simplified ANOVA equation involving genotype are of particular biological interest, because they characterize how abundance and the response to a N source and/or light is changed by the mutation or overexpression. These include: main effects of genotype, mutant and OX; pairwise interactions involving genotype, e.g. mutant:N; and 3-way interactions involving genotype, e.g. OX:Gln:L. As an example, a final gene model that includes only an intercept, an N term with positive coefficient, and and N:OX term with a negative coefficient is interpreted as a gene that is up-regulated by N in WT and mutant, less responsive to N in OX, and unregulated by light.

In *stochastic gradient descent*, we again try to obtain a linear equation for each gene or amino acid (g) as a function of a weight vector (w) and levels of condition variables (x: NO3- level, Gln, NO3\*Asn etc.). The learning problem is to determine the weights. We start with a random weight vector, then each iteration of the gradient descent forms the dot product of the weight vector with the levels of the condition variables. If the result overestimates the size of the genes, then the weights are adjusted in such a way as to reduce the overestimate (and conversely for underestimates). We use bootstrapping to determine a coefficient of adjustment (to avoid oscillations in the weights) and an L1 regularization term to eliminate small weights. Thus the cost function is:

**Feasibility, outcomes, and limitations:** This analysis will enable us to test the hypothesis that Gln/Asn ratio regulates the expression of ASN1 gene network and to develop new hypotheses for metabolic regulation genome wide. One limitation of this Aim is cost. For GC-MS analysis of amino acid pools and labeling, we will employ the expertise of the Proteomics and Metabolomics Facility at Colorado State University, which has proven to be an efficient and cost-effective approach for our analyses in the past. Another cost issue is the number of ATH1 chips to run for the proposed Microarray experiments in Aim 3B. The entire set of possible treatments for one mutant is: 4 N treatments (KCl, KNO3-, Gln, Asn) x 2 light conditions (L vs D) x 3 replicates= 24 chips/mutant. As a first test, we will initially only run ATH1 chips for KCL vs KNO3- treatments of WT and 35S::ASN1 plants in dark conditions. This is based on our previous result that the 35S::ASN1 line has elevated levels of endogenous Asn (Lam et al, 2003). This will allow us to identify candidate genes showing an altered response to endogenous Asn. We can then use these genes to perform qPCR analysis on the other mutants/treatment combinations to select which mutants/transgenics to select for the full analysis for gradient descent modeling.