Keywords: transient expression, design of experiments, plant-produced biopharmaceuticals, model building, recombinant antibodies
Abbreviations: C.V., coefficient of variation; DoE, design of experiments; dpi, days post injection; dps, days post seeding; FDA, Food and Drug Administration; Pos, position on a leaf

Plants have been developed as an alternative platform for the production of biopharmaceutical proteins, culminating recently with the FDA approval of the first plant-derived recombinant pharmaceutical enzyme for human use (ELELYSO® by Protalix Biotherapeutics). Among the many different plant-based technologies that have been proposed, transient expression mediated by *Agrobacterium tumefaciens* has proven to be particularly suitable for the rapid production of vaccines in response to emerging pandemics. However, one potential drawback of transient expression in whole plants is the large variation in recombinant protein expression levels among different leaves, which introduces a level of uncertainty in process design that can increase the regulatory burden and production costs. Transient expression is also used to test expression constructs prior to the longer and more expensive process of generating transgenic plants, and here the variation can produce misleading results leading to erroneous conclusions about the relative activity of different promoters and other regulatory elements. Such variation can be caused by loosely controlled environmental and process factors such as incubation temperature, plant characteristics and the method and timing of harvesting.

We have recently quantified the effect of several of these factors by compiling the data in a design of experiments (DoE) approach, thus building a model that can predict the amount of recombinant protein produced under different process conditions—and the resulting protein concentrations—on the overall production costs for a plant-derived monoclonal antibody. Here we discuss differences between transgenic plants and transient expression in intact plants, and their specific pitfalls for model building. We also highlight which aspects researchers should consider when using a DoE approach to investigate protein expression in plants, both for fundamental research and process development.

A detailed description of the methods can be found in our earlier manuscript. Briefly, *Agrobacterium tumefaciens* strain GV3101:pMP90RK was transformed with plant expression vector pGFD, a derivative of pPAM (GenBank AY027531) allowing transient expression of the fluorescent protein DsRed (GenBank AF168419; R2G mutant) and monoclonal antibody 2G12 (kappa light chain F62 version and gamma heavy chain) under the control of the *Cauliflower mosaic virus* (CaMV) 35S promoter. Bacteria were diluted with water and 2× infiltration medium to achieve the necessary OD$_{600}$nm value and were injected into tobacco leaves using a 1-mL syringe. The tobacco plants (*Nicotiana tabacum* cv Petit Havana SR1) were cultivated in a greenhouse at 25/22°C day/night temperature with a 16 h photoperiod (180 μmol s$^{-1}$ m$^{-2}$; λ = 400–700 nm) and at 70% relative humidity for either 35 or 42 d prior to infiltration. Infiltrated plants were incubated in phytotrons for 5 d at 70% relative humidity with a 16 h photoperiod (75 μmol s$^{-1}$ m$^{-2}$; λ = 400–700 nm) at 15, 17, 20, 22, 25, 28 or…

Research Note to: Buyel JF, Fischer R. Predictive models for transient protein expression in tobacco (*Nicotiana tabacum* cv) can optimize process time, yield, and downstream costs. Biotechnol Bioeng 2012; PMID:22511291; http://dx.doi.org/10.1002/bit.24523

www.landesbioscience.com Bioengineered 21
Transgenic plants tend to accumulate similar levels of recombinant proteins in all leaves, e.g., we recently observed a coefficient of variation (C.V.) of 36% across all leaves in transgenic tobacco plants producing DsRed and 2G12 (Fig. 1). Therefore, all leaves are approximately equal in value in terms of the product content, so a simple sampling procedure, such as the isolation and testing of one leaf from each plant, is sufficient to estimate the total amount of product formed in a batch and downstream equipment can be scaled accordingly. In contrast, our recent model predicted a sharp increase in protein levels between leaves 4 and 5, counting from the bottom of the plant, when transient expression is used to produce the same two recombinant proteins in plants of similar age. The prediction was confirmed empirically, resulting in a C.V. of 69% (Fig. 1). There was no sharp increase when younger plants were used, probably reflecting (1) the absence of a distinct sink-source transition between the leaves of young plants and (2) the more homogeneous capacity for protein biosynthesis in these leaves. A more complex sampling procedure is therefore required for transient expression systems in order to predict batch yields accurately.

One way to mitigate the additional sampling requirements in transient expression experiments is to normalize our recently-described model. Normalization can be achieved by dividing the amount of target protein produced in the different leaves by the value of a "standard" leaf, e.g., the leaf with the highest expression level. The values for 2G12 and DsRed can be used to accommodate proteins targeted to different subcellular compartments, because 2G12 is secreted to the apoplast whereas DsRed is imported into the plastids (Table 1). The normalized values can be multiplied with the amount of a target protein detected in the "standard" leaf. We are currently investigating whether the normalized expression values obtained for 2G12 and DsRed can be extrapolated to other proteins and are transferable between laboratories that use different cultivation practices.

An alternative way to predict accurate values for the concentrations of target proteins synthesized in different leaves by transient expression is to set up a custom-designed DoE and model, but this requires careful consideration of several characteristics of the transient expression system. We have already noted the crucial impact of leaf and plant age, but even though both parameters can adopt defined values (such as 47 d post seeding) these values do not necessarily describe the same physiological state of a plant at all times. For example, we found that seasonal variation has an impact on the growth rate and thus...
Finally, sampling times should be scheduled precisely as indicated by our latest results (manuscript in preparation) because recombinant protein levels vary substantially between 4 and 8 d post-infiltration, the typical range for harvest. The optimal temperature reduced DsRed expression by 25% and 2G12 expression by 15%, yet temperature changes of this magnitude can occur even in carefully-regulated environments. For example, we found that DsRed concentrations changed by 20% over 8 h on the fifth day post-infiltration (Fig. 3B).

Regardless of any pitfalls of transient expression in plants, reliable detection methods for physiological parameters and protein accumulation must be implemented to ensure that the DoE approach yields high-quality data. Among the various design types discussed in the literature for different scientific applications, the D-optimal and IV-optimal designs combined with response surface methods are thought to be the most suitable for the development of accurate predictive models. These can be implemented using software suites such as DesignExpert, JMP, Minitab, MODDE or STATISTICA. Important aspects of the design include the number of factors considered in the model, the mean standard error of the model required throughout the design space, and thus the number of single experiments included in the design. So-called block factors can also be relevant for the modeling of biological systems if the individual experiments within the design cannot be performed at the same time, e.g., using a single batch of plants. This is because block factors compensate in part for the natural variation associated with different batches of biological samples, thus improving the quality of the final model. The base model is another important aspect when modeling biological systems. This defines the type, maximum number and polynomial order of factor interactions that can be resolved by the final model, i.e., a linear base model (first order) will not allow the quadratic effect of a factor (second order) to be incorporated into the final model. For example, the temperature optimum we identified for the transient expression of DsRed and 2G12 in tobacco leaves required a quartic (forth order) model for the temperature factor because of the asymmetric (non-parabolic) shape of the optimum. Therefore, additional data from the literature or from initial screening experiments can help to improve and adjust the base model of a design, thus reducing or even eliminating the need for additional experiments to correctly describe the effect of all factors contributing to a specific response.

DoE approaches including model building can therefore improve our understanding of biological systems as exemplified by the transient expression of recombinant proteins in intact tobacco plants. Such models can help to evaluate the significance of observed effects in fundamental research projects, but when applied to actual processes they are useful to improve process design and scalability by providing accurate predictions of the impact of process variations on expression levels and protein recovery.

Table 1. Expression levels of recombinant proteins achieved 5 d post-infiltration, with values normalized to “standard” leaf six for plants harvested 47 d after seeding or leaf 4 for plants harvested 40 d after seeding.

| Plant age at harvest (d) | DsRed | 2G12 |
|-------------------------|-------|------|
| 47                      |       |      |
| Leaf                    | Amount per leaf [µg] | Normalized [-] | Amount per leaf [µg] | Normalized [-] | Amount per leaf [µg] | Normalized [-] |
| 1                       | 0.05  | 0.09 | 0.15 | 0.32 | 0.02 | 0.03 | 0.12 | 0.10 |
| 2                       | 0.10  | 0.18 | 0.32 | 0.67 | 0.06 | 0.09 | 0.40 | 0.34 |
| 3                       | 0.28  | 0.54 | 0.47 | 0.97 | 0.16 | 0.25 | 0.79 | 0.68 |
| 4                       | 0.28  | 0.53 | 0.48 | 0.90 | 0.20 | 0.31 | 1.17 | 1.00 |
| 5                       | 0.55  | 1.03 | 0.20 | 0.42 | 0.58 | 0.90 | 0.57 | 0.49 |
| 6                       | 0.53  | 1.00 | —    | —    | 0.64 | 1.00 | —    | —    |
| 7                       | 0.43  | 0.81 | —    | —    | 0.78 | 1.22 | —    | —    |
| 8                       | 0.27  | 0.50 | —    | —    | 0.48 | 0.74 | —    | —    |

*Plants harvested 40 d after seeding only possess five leaves, therefore no values are available for leaves 6-8.
Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Acknowledgments
The authors would like to thank Dr Richard M. Twyman for manuscript editing and Dr Jürgen Drossard for providing reference data on recombinant protein expression in transgenic plants.

References
1. Shoji Y, Farrance CE, Bautista J, Bi H, Musiychuk K, Horesh A, et al. A plant-based system for rapid production of influenza vaccine antigens. Influenza Other Respi Viruses 2012; 6:204-10; PMID:22094841; http://dx.doi.org/10.1111/j.1750-2659.2011.00295.x.
2. Fischer R, Schilling S, Hubrig J, Twyman RM, Drossard J. GMP issues for recombinant plant-derived pharmaceutical proteins. Biotechnol Adv 2012; 30:434-9; PMID:22050403; http://dx.doi.org/10.1016/j.biotechadv.2011.08.007.
3. Ma JK, Chakravarti B, Satter P, Fischer R, Mahoney E, Twyman RM. Planed-derived pharmaceuticals-the road forward. Trends Plant Sci 2011; 16:580-5; PMID:21708494; http://dx.doi.org/10.1016/j.tplants.2011.02.002.
4. Buyel JF, Fischer R. Predictive models for transient protein expression in tobacco (Nicotiana tabacum L.) can optimize process time, yield, and downstream costs. Biotechnol Bioeng 2012; PMID:22912122; http://dx.doi.org/10.1002/bit.25422.
5. Wyklonski T, Tomczak A, Michaluk J. Optimization of Agrobacterium-mediated transient assays of gene expression in tobacco, tomato and Arabidopsis. Plant Biotechnol J 2005; 3:259-75; PMID:15772435; http://dx.doi.org/10.1111/j.1467-7652.2005.00123.x.
6. Holland T, Jack M, Rahmacher T, Schade K, Altermann F, Staffelbach J, et al. Oxygen-nitrogen supply as a key to increased and sustained productivity of a monoclonal full-size antibody in HIV-1 reservoir culture. Biotechnol Bioeng 2010; 110:778-88; PMID:20504094; http://dx.doi.org/10.1002/bit.23648.
7. Wyźney M, Kozubek E, Lewandowski P. Optimization of transient Agrobacterium-mediated gene expression system in leaves of Nicotiana benthamiana. Acta Biochim Pol 2006; 53:285-98; PMID:16822986.
8. Stelivshyk YV, Szadkowska YH, Gerzeyenko IM, Berezovka MA, Kuchek NV. Comparison of several Nicotiana species as hosts for high-scale Agrobacterium-mediated transient expression. Biotechnol Bioeng 2007; 96:608-14; PMID:16983497; http://dx.doi.org/10.1002/bit.22185.
9. Montgomery DC. Design and Analysis of Experiments. John Wiley & Sons Incorporated, 2007.