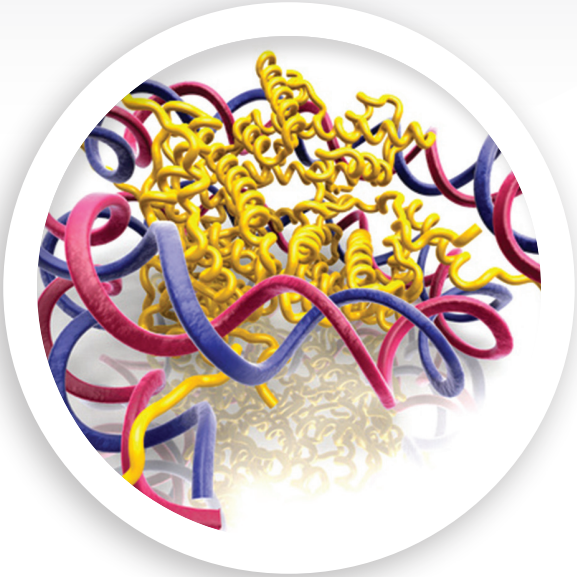
Cambridge Isotope Laboratories, Inc.
www.isotope.com

APPLICATION NOTE 20

Effective Site-Specific Isotopic Labeling (^{15}N, ^{13}C Glycine; ^{15}N, ^{13}C Phenylalanine; ^{15}N Tryptophan) Expression Optimization Using BioExpress® 2000 Media

Kenneth C. Bonanno
Vertex Pharmaceuticals, Inc.
130 Waverly Street, Cambridge, MA 02139 USA



Recombinant genes have been expressed in a variety of cellular systems for decades to generate protein reagents that are the potential targets for new small-molecule drugs. As these targets become more complex, researchers have developed innovative methods to study the structure of these proteins and the interactions with potential drugs. Cambridge Isotope Laboratories, Inc.'s BioExpress® 2000 media incorporates isotope-labeled amino acids into recombinant protein expressed in baculovirus-infected insect cells to assist NMR structural studies (Strauss, et al., 2006). To use this labeled media most efficiently, expression conditions must be optimized at a small scale prior to production.

The goal of this study was to maximize the expression of a histidine-tagged, human serine/threonine kinase in the *Spodoptera frugiperda* (SF9) cell line with site-specific isotopic labeling (^{15}N, ^{13}C-labeled glycine; ^{15}N, ^{13}C-labeled phenylalanine; ^{15}N-labeled tryptophan) in BioExpress® 2000 media (CIL #CGM-2000-Custom). The target enzyme was expressed in baculovirus-infected insect cells cultured in various small volume (2 mL) conditions. Expression was optimized by adapting a previously designed high-throughput small-scale expression protocol (Chambers, et al., 2004) and incorporating Design of Experiment (DOE, Fisher, RA, 1925).

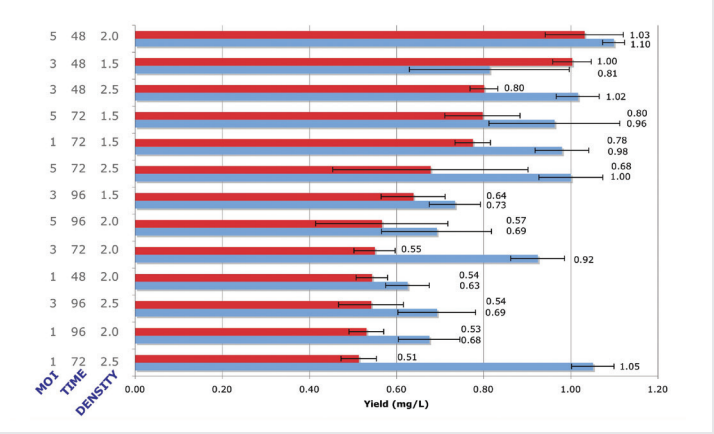
Three conditions most widely optimized for insect cell expression are:

- multiplicity of infection (MOI)
- harvest time (hours post-infection – HPI
- cell density at infection (cells/mL)

A matrix of these variables was explored in this experiment to determine the maximum protein yield, measured in mg/L.

A Box-Behnken response surface model (Box & Behnken, 1960) was applied in the experimental design using three equally spaced factors (low, medium, high) for each of the three conditions (MOI, HPI, cell density) in order to study the quantitative response, protein yield. This model provides 13 conditions, including a centerpoint with replicates. DOE allows the exploration of many variables via a limited number of experimental conditions using a statistically significant model produced by the experimental data. (Chambers and Swalley, 2008).

(continued)



Expression data of a serine/threonine kinase: labeled vs. unlabeled

Figure 1. Site-specific isotopic labeling in BioExpress® 2000 media (red) and unlabeled expression in a standard media (blue, ESF921, Expression Systems, CA). Student's t-test indicates BioExpress® 2000 media provides comparable expression yields to those of a standard media at 69% of the conditions used in the DOE experiment. (Box, Hunter, Hunter, 1978).

Cells were grown in 24 well block microplates (Whatman plc, UK) and infected with baculovirus at cell density and MOI specified by the Box-Behnken design. Cell cultures were pelleted and frozen upon harvest. Cell lysates were solubilized, and the target protein was purified using nickel magnetic agarose beads (Sigma-Aldrich, MO). Protein yield was determined using a dot blot assay with an anti-histidine primary antibody (BD BioSciences, NJ) and a goat anti-mouse secondary antibody that fluoresces at the 680 nm wavelength (LI-COR Biosciences, NE) using the Odyssey® Infrared Imaging System and the analysis software provided (LI-COR Biosciences, NE) (Figure 1).

The expression data was analyzed with statistical software (JMP™ v6.0, NC) to determine if the data sufficiently fit the model (Figure 2). The Box-Behnken model was verified since yield data provided adequate fit (R^2=80%) and there were significant regression factors in the model (p_{yield}<0.001). Additionally, the p-value for the lack of fit test is not significant (p>0.05). The data predicted optimal conditions for the large-scale expression to generate a yield of 1.3±0.1 mg/L, a 20% improvement from the best condition in the screen. As the case with many experiments incorporating DOE, the optimal condition from the model (MOI=5, HPI=48, Cell Density=1.5 x 10^6 cells/mL) corresponds to conditions not actually tested in the initial experiment. Before producing a large-scale bioreactor production batch, a 0.5 L scale expression experiment was carried out at the optimal conditions to validate the model. A 2.8 liter Fernbach flask was used for the validation experiment since it provides comparable results to a bioreactor. Standard media was used in lieu of BioExpress® 2000-labeled media for the validation since the small-scale study indicated that expression levels were comparable between the two media types at 69% of the conditions as indicated by the Student's test one-way analysis of variance (ANOVA) (Figure 1). The best experimental condition and the optimal condition predicted by the model were expressed. The target protein was purified using nickel affinity resin (Sigma-Aldrich, MO) and protein yield quantified using LC90 Lab Chip technology (Caliper Inc., MA) to validate the yields obtained in the small-scale screen. The model's predicted expression condition produced a 40% greater yield than the best experimental condition in the standard media (Figure 3).

Optimization of labeled expression in insect cells using design of experiment (DOE)

Variables Explored	Low	Mid	High
Multiplicity of Infection (MOI)	1	3	5
Cell Density (10^6 cells/ml)	1.5	2.0	2.5
Time of Infection (hrs)	48	72	96

Green represents optimal conditions

Response surface model (RSM) used to identify optimum conditions

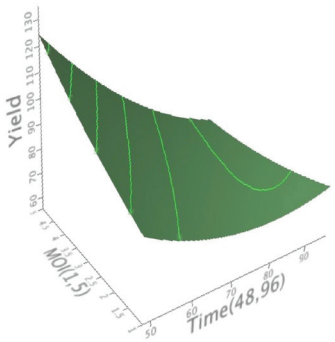


Figure 2. Insect cell expression of a recombinant human serine/threonine kinase with residue specific ^{15}N /^{13}C labeling in SF9 cells. A) A Box-Behnken design was used to maximize expression. B) The contour profile represents the RSM from the data, showing that the optimal predicted yield at cell density 1.5 x 10^6 cells/mL would be achieved at high MOI and low time.

Optimized expression condition validation

	Predicted Condition from RSD		RSD Optimal Condition
Multiplicity of Infection (MOI)	**5**		**5**
Cell Density (10^6 cells/ml)	**1.5**		**2.0**
Time of Infection (hrs)	**48**		**48**
Small Scale Yield (mg/L)	**Not Tested**		**1.1**
Validation Yield (mg/L)	**2.2**		**1.6**

Figure 3. Caliper LC90 quantitation and virtual SDS-PAGE representation of the validation expression experiment. The DOE predicted condition produced a 40% greater yield than the best condition from the model and a 100% improvement to the RSD optimal yield from the small-scale experiment.

Expression strategy for labeling in insect cells

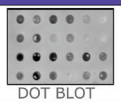
Expression Scale	Volume (ml)	Vessel	Media	Yield (mg/L)	Analysis
DOE Screening	2	Deep Well Block	Cambridge Isotopes BioExpress 2000	0.5 - 1.0	 DOT BLOT
Validation	500	Shake-Flask	Expression Systems ESF921	2.2	 SDS-PAGE
Production	2000	Wave-Reactor	Cambridge Isotopes BioExpress 2000	1.3	 SDS-PAGE

Figure 4. Expression summary. DOE screening occurred in a high throughput format and was analyzed using infrared fluorescence dot blot with Li-Cor quantitation. Validation and production yields were analyzed and quantified with Caliper LC90, virtual SDS-PAGE shown above. Validation confirmed the optimal condition defined by the DOE experiment, using unlabeled medium. Large-scale expression was then performed using the optimal condition defined by the DOE experiment using labeled medium.

The hypothesized optimal expression conditions (MOI = 5, HPI = 48, cell density = 1.5 x 10^6 cells/mL) were used for production in the BioExpress® 2000 media using protocols recommended by CIL. One liter of cells was cultured in a 2 L Wave Bioreactor bag for each batch (GE Healthcare, NJ). The final yield of the affinity-purified protein was 1.3 mg/L, within error of the model's prediction for labeled media. Figure 4 describes the process implemented from screening through validation into production.

It is important to connect laboratory technique and procedures with statistical methods to maximize the experimental efficiency at every level of experimentation. Incorporating isotopically labeled amino acids into insect cell media increases the cost of the media; therefore, optimal expression conditions for the target protein are desired before advancing to this media. DOE and statistical analysis control the size of the experiment and allow efficient use of reagents, such as labeled media, for optimization purposes. This enables more of the labeled media to be used for the scale-up processes and ultimately to supply material to generate NMR samples. Designing a screening experiment using high throughput miniaturization and sound statistical models minimizes reagent consumption while maximizing the expression yields of protein production batches.

Abbreviations

ANOVA: One-way analysis of variance
DOE: Design of experiment
HPI: Hours post infection
MOI: Multiplicity of infection
RSD: Response surface design
RSM: Response surface model
SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SF9: *Spodoptera frugiperda*

Acknowledgements

I thank Ms. Judith Lippke and Ms. Fan Lu who offered their technical expertise in cloning and generating the baculovirus used in this project. Mr. Douglas Austen provided guidance and advice throughout the experiments and data analysis. I thank Mr. John Fulghum and Drs. Susanne Swalley, Peter Connolly, Marc Jacobs and Ted Fox for reviewing the manuscript. Special thanks to Dr. Lynne Batchelder, Dr. Kevin Millis and Mr. Gerry Guilfoy for their interest and support in pursuing this manuscript at Cambridge Isotope Laboratories, Inc.

(continued)

Additional products of interest

CGM-2000-U	BioExpress® 2000 (unlabeled)
CGM-2000-U-s	BioExpress® 2000 (unlabeled) 200 mL media kit
CGM-2000-N	BioExpress® (U-^{15}N, 98%)
CGM 2000-CN	BioExpress® (U-^{13}C, 98%; U-^{15}N, 98%)
DLM-2622	DL-1,4-dithiothreitol (DTT) (D_{10}, 98%)
DLM-2274	Dodecylphophocholine (DPC) (D_{38}, 98%)
CDLM-7572	L-α-phosphatidylcholine, dipalmitoyl (DPPC) (palmitate-U-$^{13}C_{32}$, 98%+; methyl-D_9 choline, 98%)
DLM-197	Sodium dodecyl sulfate (SDS) (D_{29}, 98%)
DLM-1814	Tris(hydroxymethyl) methylamine (TRIS) (D_{11}, 98%)
DLM-3033	Imidazole (D_4, 98%)
DLM-3786	N-(2-Hydroxyethyl) piperazine-N′-2-ethanesulfonic acid (HEPES) (D_{18}, 98%)
CNLM-2408-1	GFL peptide standard (^{13}C, 98%; ^{15}N, 96-99%)
CNLM-6840-10	SH3 domain protein (U-^{13}C, 98%; U-^{15}N, 98%)
NLM-7361-5	Ubiquitin (algal) (U-^{15}N, 98%)
CNLM-7362-5	Ubiquitin (algal) (U-^{13}C, 98%; U-^{15}N, 98%)

References

Box, G.; Behnken, D. **1960**. Some new three level designs for the study of quantitative variables, *Technometrics, 2,* 455-475.

Chambers, S.P.; Austen, D.A.; Fulghum, J.R.; Kim, W.M. **2004**. High-throughput screening for soluble recombinant expressed kinases in *Escherichia coli* and insect cells. *Protein Expression and Purification, 36,* 40-47.

Strauss, A.; Fendrich, G.; Jahnke, W. **2006**. Effective uniform isotope labeling of proteins expressed in baculovirus-infected insect cells using BioExpress® 2000 (insect cell) medium. Cambridge Isotope Laboratories Application Note, *14,* 1-3.

Chambers, S.P.; Swalley, S. **2008**. Designing Experiments for High-Throughput Protein Expression. High Throughput Protein Expression and Purification: *Methods in Molecular Biology*. Ed. Dr. Sharon Doyle. (in press).

Fisher, R.A. **1925**. Statistical Methods for Research Workers (ISBN 0-05-002170-2).

Box, G.; Hunter, W.G.; Hunter, J.S. **1978**. Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building (ISBN 0-471-09315-7)