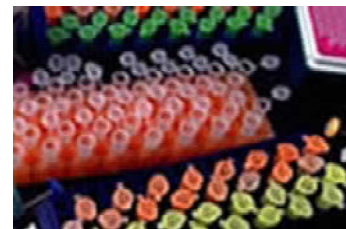


## cDNA Cloning

PRINT

The cDNA CLONING Team, headed by Brian Fox, PhD, is responsible for obtaining and manipulating cDNAs in advance of protein production. This includes the production of vectors designed for this purpose.



### Goals for CESG cDNA Cloning Team

- Construct up to 200 expression clones per month from the list of ORFs selected by the Bioinformatics Team.
- Determine the DNA sequence of the clones and enter any differences into the database.
- Develop expression vectors and other vector resources for the project.

- [Bioinformatics](#)
- [Cell-Free](#)
- [Cloning](#)
- [Crystallography](#)
- [Expression Testing](#)
- [NMR](#)
- [Protein Production](#)
- [Protein Purification](#)
- [Quality Assurance](#)
- [Sesame](#)

### Overview of cDNA Cloning at CESG

The success of structural genomics initiatives is critically dependent on the incorporation of targeted open reading frames (ORFs) into vectors suitable for protein production. Among the many structural genomics groups using Gateway® recombination, CESG previously developed and used the Gateway® method to clone ~3500 eukaryotic ORFs. As part of this effort, a customized modular vector backbone was created to allow efficient swap of antibiotic resistance markers, protein tags, linker regions, and protease sites. Some of the best variants of these production vectors, available by material transfer agreement with the University of Wisconsin, are described.

To facilitate expansion of this vector set to include wheat germ cell-free and other expression platforms, we evaluated Flexi®Vector, a restriction enzyme/ligation based cloning system recently developed by [Promega Corporation](#) (Madison, WI). This system offered the advantages of high-throughput cloning of PCR products directly into an expression vector and serial transfer of the sequenced verified ORFs from the first vector to others. Here we report a comparison of Gateway® recombination cloning system and the Flexi®Vector restriction-based cloning system.

Cloning protocols for each system were conducted in parallel for 96 different target genes from PCR through the production of sequence verified expression clones. The shorter nucleotide sequences required to prepare the target ORFs for Flexi®Vector cloning allowed a single-step PCR protocol, resulting in fewer mutations relative to the Gateway® protocol. Furthermore, through initial cloning of the target ORFs directly into an expression vector, the Flexi®Vector system gave time and cost savings compared to the CESG protocol originally developed for the Gateway® system. Within the Flexi®Vector system, genes were transferred between four different expression vectors. The efficiency of gene transfer between Flexi®Vectors depended on including a region of sequence identity adjacent to one of the restriction sites. With the proper construction in the flanking sequence of the vector, gene transfer efficiencies of 95-98% were obtained. Detailed protocols developed for the Flexi®Vector method, the current catalog of vectors developed for this project, and opportunities for multiplexed cloning and expression studies are presented.

There are strengths and weaknesses inherent to any cloning system. The cloning steps in the Gateway® system are highly efficient, and there are a wide variety of vectors available due to the length of time this system has been in use. However, the requirement for an initial, non-productive cloning step and the long primer sequences required to encode the recombination sites are drawbacks to this system. In the Flexi®Vector system, the initial cloning step inserts the target gene into an expression vector and the short flanking nucleotide sequences can be added in a single PCR step. As demonstrated here, the efficiency of Flexi®Vector cloning can match or exceed that of recombination cloning for both initial capture of PCR products and for transfer between different vector combinations. These advantages lead to savings in time

## Center for Eukaryotic Structural Genomics - cDNA Cloning

and cost, and fewer mutations present in the expression clones. There currently fewer vector options available for the Flexi@Vector system than for Gateway@ and target genes must be screened for the presence of the Sgf I and Pme I sites.

We conduct a high-throughput Plasmid DNA screen to determine the presence of insert in both our Entry and Expression plasmids. A robot-aided protocol was developed whereby the colonies are picked into 96-well growth blocks containing CircleGrow™ media and the appropriate antibiotic. This is grown overnight at 37°C with vigorous shaking. The next morning plasmid DNA is isolated with the use of QiaRobot 8000. This DNA is then used as a template in PCR using universal vector primers that flank the insertion site. The PCR products are analyzed on E-gels 96 (Invitrogen) and positive clones moved down the pipeline.

## Methods Being Investigated and Publications

### + Overview of High-Throughput Cloning Protocol

#### A. Work Group Design

1. Selection of 96 ORFs.
2. Automatic primer design.

#### B. ORF Amplification

1. Two-step PCR from T87 cDNA.
2. PEG precipitation.

#### C. Entry Vector Cloning

1. Gateway@ BP clonase reaction.
2. 96-well transformation into Top10.
3. HTP plating using ColiRollers glass beads.

#### D. Entry Vector Colony Screening

1. Pick two colonies per transformation.
2. Inoculate into 96-well block containing CircleGrow media.
3. Grow overnight.
4. Mini plasmid prep using QiaRobot 8000.
5. PCR screen using M13 forward and reverse primers.
6. Analyze on E-gel 96.

#### E. Sequencing

1. Pick one positive clone based on screen.
2. Amplify template using TempliPhi.
3. Fluorescent sequencing reactions.
4. Gel run at UW-Madison Biotech Center.
5. Analyze sequence data.
6. Store data that differs from MIPS database, discard clones that don't contain complete reading frame.
7. Rearray sequence positive entry clone DNA.

#### F. Destination (Expression) Vector Cloning

1. Gateway@ LR clonase reaction into pVP13-GW.
2. 96-well transformation into Top10.
3. HTP plating using ColiRollers glass beads.

#### G. Destination Vector Colony Screening

1. Pick two colonies per transformation.

2. Inoculate into 96-well block containing CircleGrow media.
3. Grow overnight.
4. Mini plasmid prep using QiaRobot 8000.
5. PCR screen using MBP forward and pQE reverse primers.
6. Analyze on E-gel 96.
7. Rearray PCR positive destination clone DNA.

#### + Structural Genomics Methods Applied to Production of TEV Protease

Tobacco etch virus N1a proteinase (TEV protease) is an important tool for the removal of fusion tags from recombinant proteins. Production of TEV protease in *E. coli* has been hampered by insolubility and addressed by many different strategies. However, the best previous results and newer approaches for protein expression have not been combined to test whether further improvements are possible. Here we use a quantitative, high throughput assay for TEV protease activity in cell lysates to evaluate the efficacy of combining several previous modifications with new expression hosts and induction methods. Small-scale screening, purification and mass spectral analysis showed that TEV protease with a C-terminal poly-Arg tag was proteolyzed in the cell to remove 4 of the 5 arginine residues. The truncated form was active and soluble but in contrast, the tagged version was also active but considerably less soluble. An engineered TEV protease lacking the C-terminal residues 238-242 was then used for further expression optimization. From this work, expression of TEV protease at high levels and with high solubility was obtained by using auto-induction medium at 37°C. In combination with the expression work, an automated two-step purification protocol was developed that yielded His-tagged TEV protease with >99% purity, high catalytic activity and purified yields of ~400 mg/L of expression culture (~15 mg pure TEV protease per g of *E. coli* cell paste). Methods for producing glutathione S-transferase tagged TEV with similar yields (~12 mg pure protease per g of *E. coli* cell paste) are also reported. These vectors are available by completion of standard biological materials transfer agreement, and have been deposited in the NIH [PSI-Materials Repository \(PSI-MR\)](#).

Blommel, P.G., Fox, B.G. (2007) **A combined approach to improving large-scale production of tobacco etch virus protease.** *Protein Expr Purif* 55(1):53-68. [[17543538](#)]

#### + Structural Genomics Methods Applied to Production of Monotopic Membrane Protein

Fusion protein vectors developed for high-throughput protein expression as part of the Protein Structure Initiative have been investigated for use in the expression and stabilization of human cyt b5, a monotopic membrane protein that must be attached to the cellular membrane for function. Expression as a fusion to His8-maltose binding protein allowed expression of the full-length cyt b5 as a fully soluble entity. Maintenance of the solubility in *E. coli* during the time course of expression was associated with high-level incorporation of protoporphyrin IX into the heme domain of the fusion protein. The full-length cyt b5 could be liberated from the fusion by site-specific proteolysis, which permitted spontaneous incorporation into membrane vesicles. This work provides a method for the production and high-yield in situ delivery of monotopic membrane proteins to lipid environments. Also funded by NIH GM 50853, B.G. Fox, PI.

Sobrado, P., Goren, M.A., James, D., Amundson, C.K., Fox, B.G. (2008) **A Protein Structure Initiative approach to expression, purification, and *in situ* delivery of monotopic membrane protein human cytochrome b5 to membrane vesicles.** *Protein Expr Purif* 58(2):229-41. [[18226920](#)]

#### + Customized Expression Vector Platform

A series of expression vectors used for pipeline protein production and different research activities are available. These vectors use Gateway®, Flexi®iVector, or restriction digestion cloning methods. The vector backbones have been modularized for simple exchange of promoters, affinity tags and solubility tags. Over seventy versions are currently available and support expression in bacteria, wheat germ lysates, and insect lysates. Sequence verified genes can be transferred between each of these expression platforms by simple, highly efficient methods. These vectors are available by completion of standard biological materials transfer agreement, and have been deposited in the [NIH PSI-Materials Repository](#).

Blommel, P.G., Fox, B.G. (2007) **A combined approach to improving large-scale production of tobacco etch virus protease.** *Protein Expr Purif* 55(1):53-68. [[17543538](#)]

Blommel, P.G., Fox, B.G. (2005) **Fluorescence anisotropy assay for proteolysis of specifically labeled fusion proteins.** *Anal Biochem* 336(1):75-86. [[15582561](#)]

Blommel, P.G., Martin, P. A., Seder, K. D., Wrobel, R.L., Fox, B.G. (2007) **Flexi®iVector Cloning**, Methods in Molecular Biology, J.E. White, Editor. *Humana Press*, Totowa, NJ.

Blommel, P.G., Martin, P.A., Wrobel, R.L., Steffen, E., Fox, B.G. (2006) **High-efficiency single-step production of expression plasmids from cDNA clones using the Flexi®Vector cloning system.** *Protein Expr Purif* 47(2):562-70. [16377204]

Frederick, R.O., Bergeman, L., Blommel, P.G., Bailey, L.J., Song, J., Meske, L., Bingman, C.A., Ritters, M., Dillon, N., Kunert, J., Yoon, J., Lim, A.-Y., Cassidy, M., Bunge, J., Aceti, D.J., Primm, J.P., Markley, J.L., Phillips, G.N., Jr., Fox, B.G. (2007) **Small-scale, semi-automated purification of eukaryotic proteins for structure determination.** *JSFG* 8(4):153-66. [17985212]

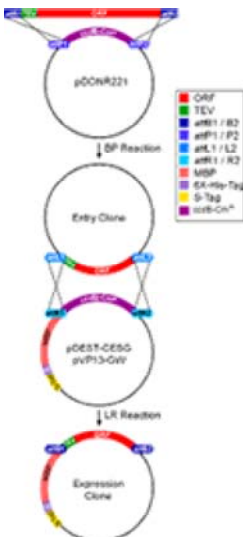
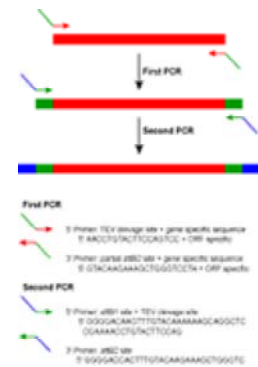
Thao, S., Zhao, Q. Kimball, T., Steffen, E., Blommel, P. G., Ritters, M., Newman, C. S., Fox, B. G., Wrobel, R.L. (2004) **Results from high-throughput DNA cloning of *Arabidopsis thaliana* target genes using site-specific recombination.** *JSFG* 5(4):267-76. [15750721]

**+ Gateway®Cloning at CESG**

CESG uses Gateway® Technology (Invitrogen) to generate expression vectors that provide simple swapping of expression systems and protein tags. CESG has developed expression vectors that add S-tag (for detection), a His6-tag (for purification), and MBP (maltose-binding protein, for solubilization and purification) to the N-terminal of the target protein. When required, the entire fusion is cleavable from the protein target by TEV protease. These vectors are derived from pET (T7 promoter; Novagen), pQE (T5 promoter; Qiagen), and pBAD (AraBAD promoter; Invitrogen) vector backbones.

We use a two-step PCR protocol to amplify ORFs from our cDNA pool and to incorporate the recombination (att) and TEV protease cleavage sites.

**Strategy for Generation of attB-PCR Substrates for Recombinational Cloning:** The figure at left depicts the two-step polymerase chain reaction (PCR) amplification strategy developed by the CESG for testing the Gateway® Recombination cloning technology. The first PCR makes use of a 5' primer that incorporates a portion of the TEV protease cleavage site immediately prior to the start of the target ORF, and a 3' primer that incorporates a recombination sequence distal to the stop of the ORF. The second PCR proceeds with 5'- and 3'- primers that generate the complete attB1-TEV and attB2 sites, producing the attB-PCR substrate, attB1-TEV-ORF-attB2, that is suitable for recombination. (Click on the image at right to see larger sized-image.)



**Generation of Entry and Destination Clones Using Gateway Recombinational Cloning Technology:** The figure at left is a schematic representation of the recombination cloning method based on the site-specific recombination properties of bacteriophage  $\lambda$ . A polymerase chain reaction (PCR) product flanked by attB sites and consisting of the TEV protease cleavage site prior to the start of the target gene, attB1-TEV-ORF-attB2, is transferred to donor vector pDONR221, that contains the cassette attP1-ccdB-attP2 by way of a BP recombination reaction, producing an attL1-TEV-ORF-attL2 entry clone following  $\lambda$  integration and recombination. The attL-PCR insert is transferred to an attR substrate, our pDEST-CESG destination vector, pVP13-GW, created by the CESG to add an S-Tag, 6XHis-Tag, maltose binding protein (MBP) fusion to the N-terminus of recombinant proteins that is cleavable by TEV protease that has been converted to accommodate recombinational cloning, by way of an LR *in vitro* recombination reaction, regenerating the attB sites by producing an attB1-TEV-ORF-attB2 expression clone. (Click on the image at left to see larger sized-image.)

**+ Example of Two-Step PCR Amplification of cDNA Derived from *Arabidopsis* T87 Callus Culture RNA**

The current CEGS protocol uses RNA isolated from the T87 callus culture line to produce the cDNA. Using this RNA source we are able to amplify about 60-80% of the targeted ORFs. This success rate reflects our analysis of gene chips produced by NimbleGen's maskless array DNA synthesis technology. This chip analysis showed that about 60-80% of all *Arabidopsis* ORFs are expressed in this callus cell line. These chip results are being integrated into Genie to optimize workgroup generation and to guide researchers in cDNA target selection. The T87 cell culture was kindly provided by Drs. B.H. Kang, D. Rancour, and S. Bednarek.

• Distribution of Biological Materials via the PSI Materials Depository (PSI-MD)

## Center for Eukaryotic Structural Genomics - cDNA Cloning

### [↑ Distribution of Biological materials via the PSI Materials Repository \(PSI-MR\)](#)

The Protein Structure Initiative (PSI), funded by the US National Institutes of Health (NIH), provides a framework for the development and systematic evaluation of methods to solve protein structures. Although the PSI and other structural genomics efforts around the world have led to the solution of many new protein structures as well as the development of new methods, methodological bottlenecks still exist and are being addressed in this 'production phase' of PSI. The PSI has charged all centers to participate in a plan to "store and distribute PSI clones generated by the PSI Centers." While the [PSI-Material Repository \(PSI-MR\)](#) is a separately funded activity, CESG has allocated significant resources in developing a plan to fulfill the important mission of transferring clones and plasmids to the PSI-MR. Issues ranging from intellectual property rights to tactical sample handling, materials verification, inter-site communication protocols, and permanent data storage have been addressed to facilitate the transfer, verification, tracking, and distribution of expression clones and other materials. A preliminary success rate for transfer of eukaryotic expression clones has been established, and other PSI-derived expression vectors and materials have been deposited and are becoming available for distribution. Although our center incurred substantial costs in establishing this transfer pipeline, we foresee that we will benefit from the long-term savings resulting from eliminating the need for the individual, customized materials transfer agreements and internal preparations that were previously required prior to shipping CESG's clones and vectors to requestors around the globe. Ongoing challenges associated with transfer of larger numbers of expression clones will need to be addressed.

Fox, B.G., Goulding, C., Malkowski, M.G., Stewart, L., Deacon, A. (2008) **Structural genomics: from genes to structures with valuable materials and many questions in between.** *Nat Methods* 5(2):129-32. [[18235432](#)]

## cDNA Cloning Team

- Brian Fox, PhD (PI)
- Russell Wrobel, PhD

---

Center for Eukaryotic Structural Genomics (CESG), Department of Biochemistry  
University of Wisconsin-Madison, 445 Henry Mall, Madison, WI 53706 | [Map](#) |  
Telephone: 608.263.2183 Fax: 608.890.1942 Email: [cesginfo@biochem.wisc.edu](mailto:cesginfo@biochem.wisc.edu)

| [Biochemistry](#) | [Private](#) | [PDB](#) | [PSI](#) | [PSI-KB](#) | [PSI-MR](#) | [Sesame](#) | [UW-Madison](#) |  
Copyright © 2006 The Board of Regents of the University of Wisconsin System