

# Center for Eukaryotic Structural Genomics

## Protein Structure Initiative



### Use of Cell-Free Protein Production Platform for X-ray Crystallography

Phillips, G.N., Jr., Aceti, D.J., Bingman, C.A., Burgie, E.S., Fox, B.G., Frederick, R.O., Makino, S., Markley, J.L., Primm, J.P., Vojtki, F.C., Wrobel, R.L.  
 University of Wisconsin-Madison, Department of Biochemistry, 433 Babcock Drive, Madison, Wisconsin, USA <http://www.uwstructuralgenomics.org>

#### Abstract

The Center for Eukaryotic Structural Genomics (CESG) has been working to further develop our *in vitro* transcription-wheat germ cell-free translation system for the production of proteins for structural studies. To date, sixteen NMR structures have been determined at CESG from protein produced by our cell-free platform. More recently, attention has been turned to the use of this platform for X-ray crystallography. The first structures from CESG to be determined from cell-free sources are two crystal forms of arginine iminohydrolase, one in complex with a ligand, thereby extending earlier work on this protein by CESG, and the structure of a complex of two human proteins involved in DNA replication produced by simultaneous translation in the cell-free system. The creation of this complex represents a control experiment, following earlier work on a similar complex by Deng et al.(2) Two other novel target proteins, human cystatin-M and a mammalian DCN1-like protein 2 homolog from the thermophilic algae *Galdieria sulphuraria* have been crystallized. The DCN1 homolog crystals diffract to a resolution of 1.2 Å and its structure determination has recently been completed. Crystals of cystatin-M are in our diffraction screening process. Incorporation of selenomethionine into proteins produced with the cell-free methods is as straightforward as substituting selenomethionine for methionine in the reaction mixture. There are no toxic side effects of selenomethionine, and incorporation is essentially 100 percent. These results clearly establish that the method is capable of providing support for structural genomics and structural biology research projects.

CESG's partnership with CellFree Sciences Co., Ltd. (Yokohama, Japan) and with Prof. Yaeta Endo has steadily improved the capabilities of wheat-germ based protein production systems during PSI-1 and PSI-2 (1). The costs can be made quite reasonable. While the reactions can be carried out on the bench top, robotic support allows for high-throughput screening and also improves reproducibility due to standardization of liquid handling and timing of events (Table 1). Small-scale screening on the 1-5 micromolar scale can be performed on the GenDecoder instrument for a cost of about \$1/sample (not including DNA production costs). At the next scale-up level, the procedure from DNA to purified protein can be carried out in automated fashion on the Protomist DT-II bench-top robot in sufficient quantities for small-scale crystallization screening in favorable circumstances when all channels are dedicated to one protein. Scaling to the next level, the Protomist series and Protomist XE are capable of larger scale productions but without the protein purification capability.

#### Results and Discussion

CESG has begun to use cell-free translation to solve structures using X-ray crystallography, and four have been completed since April 2009. Proof-of-concept studies were carried out with arginine iminohydrolase. A purified sample of 77 mg was obtained from 6 mL of the cell-free extract 2240H (1.3 mg/ml extract, a better than average yield for cell-free translation; our cost \$2700). For comparison, 110 mg was obtained from 2000 mL of *E. coli* culture, a considerably better than average

Table 1. Cell-Free Protein Production Robotics capabilities available to the NP Partnership.

	GenDecoder1000	Protomist DT-II	Protomist100	ProtomistXE (prototype)
<b>Typical Use</b>	HTP Screening	Screening, Characterization	Characterization, Production	Production
<b>Format</b>	96 well; 4x96 well	24 well; 6 well	8 tubes	1 reaction
<b>Total Reaction Volume</b>	14.4-19.2 mL	33.6 mL, 24 mL	32 mL	10-20 mL
<b>Reaction Time</b>	24 h	14 to 24 h	18 h	12 to 48 h
<b>Automated Transcription*</b>	Yes	Yes	Yes	No
<b>Automated Translation</b>	Yes	Yes	Yes	Yes
<b>Automated Purification*</b>	No	Yes	No	No
<b>Yield Per Reaction*</b>	1 to 5 µg	50 µg to 1 mg	0.5 to 10 mg	1 to 70 mg

\*Reaction initiated by provision of purified plasmid DNA. \*Protomist-DTI provides automated purification in the 6 well mode. \*Based on CESG results with expressed proteins. 70% of proteins tested fall inside these expression yield thresholds.

yield for a eukaryotic protein in *E. coli*). The selenomethionyl incorporation determined by mass spectrometry was 99% in the cell-free sample, exactly matching the composition of the added amino acid. The deposited structure of the native protein (PDB 3H7C, see below) from cell-free translation had a resolution of 1.5 Å, comparable to the cell-based structure (PDB 1VKP). A functional follow-up study using the same batch of cell-free protein yielded a structure of a covalent intermediate (PDB 3H7K, see below). As another test, CESG chose a heterodimer complex originally prepared by other from *E. coli* with a yield of ~0.02 mg/mL of culture. Their preparation required a complicated purification process; some crystals took as long as 1 year to grow; and the structure was solved to 2.5 Å resolution (2). In our cell-free translation system, the intact protein complex was obtained by simply adding two mRNAs to the translation reaction, with a protein yield of 2.5 mg from 10 mL of extract (WEPRO 2210H, \$4500).

The Sesame LIMS shows that the total elapsed time from availability of the two cloned genes to solved structure was 137 days. Crystals were obtained in different conditions than the previous work and had a different space group. The structure obtained from cell-free translation (PDB 3KDF, see below) had a better resolution of 2 Å. An important feature of this work was that the open nature of cell-free translation was successfully exploited to assemble target solved by X-ray crystallography. A third PDB entry, 3KEV, represents our first *de novo* cell-free structure (see below for details). The Protomist XE was used with 5 mL of WEPRO 8240H extract (our cost \$4000), and the yield of purified protein was 4.2 mg. The cell-free capability has also been used in functional studies of the human desaturase-cytochrome complex, and is incorporated into our proposed combinatorial discovery efforts with other membrane proteins. The system has also been used to produce other membrane proteins using liposomes as the initial repository of the polypeptide chains (3). As a control for the structure determination process, bacteriorhodopsin has been produced in milligram quantities and crystallizations are underway following earlier published work by others.

Advantages of the system are that it sometimes produces soluble protein when other systems fail and that it involves lower labor costs. (Fifteen project personnel were involved in the *E. coli* effort vs. eight project personnel are involved in the cell-free effort.) Disadvantages are a steep learning curve and the supply cost. The key determinant of the supply cost of this cell free system is the amount of wheat germ extract committed to a trial. As in other expression systems, many times proteins are expressed only in insoluble forms, and small-scale screening trials are advised. When producing soluble proteins, the yield per mL of wheat germ extract is variable, but an upper limit (for GFP) is about 4 mg per mL of extract in the Protomist 100. More typical numbers when the process works are 0.1-0.5 mg of purified protein per mL of extract. In the Protomist XE, which has been optimized for larger scale production using a more stringently prepared wheat germ extract, the yield per mL of extract is increased four-fold.

#### References

1. Tyler et al. Comparison of cell-based and cell-free protocols for producing target proteins from the *Arabidopsis thaliana* genome for structural studies. *Proteins* 2005 39:633.
2. Deng et al. Structure of the full-length human RPAA14/32 complex gives insights into the mechanism of DNA binding and complex formation. *JMB* 2007, 474:865.
3. Goren, et al. Cell-free translation of integral membrane proteins into unilamellar liposomes. *Methods Enz* 2009, 463:647.

Note: Emerald BioSystems is now the US distributor of CellFree Sciences Products. (Disclosure: No financial consideration has been received by the authors or CESG from Emerald BioSystems. A collaboration with CellFree Sciences Co., Ltd. has provided instruments at discounted prices and wheat germ extract and other supplies to CESG at substantial discount; authors have received no financial remuneration.)

**Center for Eukaryotic Structural Genomics**

**Protein Structure Initiative**

**Target ID:** GO:24674  
**Source Organism:** *Arabidopsis thaliana*  
**Target Name:** A5G08170.1  
**PDB Entry:** 3H7C  
**Function:** Arginine iminohydrolase (EC:3.5.3.12)  
**Produced From Structure by X-ray:** Wheat germ cell-free synthesis  
**Data Collected At:** No. of Residues: 363  
**Authors:** E.S. Burgie, C.A. Bingman, G.N. Phillips, Jr.

**Structural Features:**  
 This target was originally nominated by David Makino of Oklahoma State University and was the first externally nominated target solved by X-ray crystallography. It was selected based on an interesting phenotype in *Arabidopsis* seedlings, identified by genetic screens. Subsequent biochemical characterization by Protrowick and coworkers established A5G08170.1 as arginine iminohydrolase, the final uncharacterized step in polyamine biosynthesis in plants. This structure represents the first X-ray structure determined by CESG using protein produced by cell-free synthesis, and served as a process control. It was produced using selenomethionine in the reaction mixture and solved *de novo* using MAD phasing. The crystal form chosen is different from the original form that was solved using *E. coli* production technology and difficult to higher resolution.  
 References: (1) Janowitz, T., Kneifel, H., Protrowick, M. (2003) Identification and characterization of plant arginine iminohydrolase, the last missing link in polyamine biosynthesis of plants. *FEBS Lett* 544(1-3):258-61  
 Functional follow-up of E-1: PDB\_3zph

**Percent Identity with Nearest PDB Structure at Time Solved:** 26.12

**Protonet Cluster Size: Structures in PDB:** 2612

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.890.1942; email: [cesg@biochem.wisc.edu](mailto:cesg@biochem.wisc.edu); website: <http://www.uwstructuralgenomics.org>. This research funded by NIH/NIGMS Protein Structure Initiative grants USA GM074901 and P50 GM064598.

**Center for Eukaryotic Structural Genomics**

**Protein Structure Initiative**

**Target ID:** GO:24674  
**Source Organism:** *Arabidopsis thaliana*  
**Target Name:** A5G08170.1  
**PDB Entry:** 3H7K  
**Function:** Arginine iminohydrolase (EC:3.5.3.12)  
**Produced From Structure by X-ray:** Wheat germ cell-free synthesis  
**Data Collected At:** No. of Residues: 363  
**Authors:** E.S. Burgie, C.A. Bingman, G.N. Phillips, Jr.

**Structural Features:**  
 This target was originally nominated by David Makino of Oklahoma State University and was the first externally nominated target solved by X-ray crystallography. It was selected based on an interesting phenotype in *Arabidopsis* seedlings, identified by genetic screens. Subsequent biochemical characterization by Protrowick and coworkers established A5G08170.1 as arginine iminohydrolase, the final uncharacterized step in polyamine biosynthesis in plants. A5G08170.1 represents the first structure of an arginine deminase, and the first representative of the PPAD family. It is a member of the same superfamily as aminotransferase, unlike the PPAD and known arginine deminase family, will allow the modeling of a large number of proteins and will facilitate the elucidation of the mechanism of this enzyme. The covalent adduct described here was prepared with protein produced by CESG's wheat germ cell-free system is possible by back reaction with the product N-carbamoylputrescine in the presence of the other product, NH<sub>3</sub> (to be added).  
 References: (1) Janowitz, T., Kneifel, H., Protrowick, M. (2003) Identification and characterization of plant arginine iminohydrolase, the last missing link in polyamine biosynthesis of plants. *FEBS Lett* 544(1-3):258-61  
 Functional follow-up of E-1: PDB\_3zph

**Percent Identity with Nearest PDB Structure at Time Solved:** 26.12

**Protonet Cluster Size: Structures in PDB:** 2612

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.890.1942; email: [cesg@biochem.wisc.edu](mailto:cesg@biochem.wisc.edu); website: <http://www.uwstructuralgenomics.org>. This research funded by NIH/NIGMS Protein Structure Initiative grants USA GM074901 and P50 GM064598.

**Center for Eukaryotic Structural Genomics**

**Protein Structure Initiative**

**Target ID:** GO:102544, GO:102545  
**Source Organism:** *Human sapiens*  
**Target Name:** Human replication protein A complex  
**PDB Entry:** 3KDF  
**Function:** DNA replication fork A complex during replication  
**Produced From Structure by X-ray:** Wheat germ cell-free synthesis  
**Data Collected At:** Resolution: 1.98 Å  
**Authors:** No. of Residues: 506ASU  
 Advanced Protein Source: GMCA CAT 23-D-D 05-Aug-2009

**Structural Features:**  
 This target was selected as part of control workgroup, designed to determine the feasibility of synthesizing protein complexes by wheat germ cell-free synthesis. As such, the most interesting feature of this structure is that the protein was synthesized co-translationally of different mRNA preparations in one cell-free reaction mixture. The primary purpose of crystallization and structure determination was validation that the cell-free reaction products could be purified as a complex, and to validate that the quality of the materials was sufficient to support phasing by SeMet anomalous diffraction. The target was based on PDB entry 1LJL.  
 Functional follow-up of E-1: PDB\_3zph

**Percent Identity with Nearest PDB Structure at Time Solved:** 15.63 (17.8%)

**Protonet Cluster Size: Structures in PDB:** 154, 3, 24, 3

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.890.1942; email: [cesg@biochem.wisc.edu](mailto:cesg@biochem.wisc.edu); website: <http://www.uwstructuralgenomics.org>. This research funded by NIH/NIGMS Protein Structure Initiative grants USA GM074901 and P50 GM064598.

**Center for Eukaryotic Structural Genomics**

**Protein Structure Initiative**

**Target ID:** GO:68344  
**Source Organism:** *Galdieria sulphuraria*  
**Target Name:** DCN1L domain containing protein  
**PDB Entry:** 3KEV  
**Function:** DCN1L domain containing protein  
**Produced From Structure by X-ray:** Wheat germ cell-free synthesis  
**Data Collected At:** Resolution: 1.84 Å  
**Authors:** No. of Residues: 198  
 Advanced Protein Source: GMCA CAT 23-D-D 12-Oct-2009  
 Advanced Protein Source: GMCA CAT 23-D-D 05-Oct-2009

**Structural Features:**  
 This target was selected as part of a pair of workgroups, designed to compare outcomes from medially relevant human proteins, and homologs from the thermophilic alga, *Galdieria sulphuraria*. In this case, the sequence of 3KEV is 50% identical to its human ortholog, whereas the identity between the closest structures of PDB to the human ortholog was less than 30%. The human ortholog listed in our standard processing, so this structure marks a success for this paired target selection methodology for determining the structures of highly conserved proteins relevant to human biology and health. Biochemical studies in mice indicate that this protein probably serves as a scaffold protein involved in protein neddylation. Protein neddylation is involved in apoptosis, and defects in neddylation are implicated in many important biological phenomena, such as cancer, Alzheimer's disease, and life-span. It is also noteworthy that this was CESG's first initial target protein to be synthesized by cell-free techniques that has progressed to the Protein Data Bank.  
 1. Gao, et al. (2006) Nature Cell Biology Oct(10):1171-7  
 2. Chen (2004) Apoptosis July 9(4):415-422

**Percent Identity with Nearest PDB Structure at Time Solved:** 38.03 (29.0%)

**Protonet Cluster Size: Structures in PDB:** 3803 (29.0%)

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.890.1942; email: [cesg@biochem.wisc.edu](mailto:cesg@biochem.wisc.edu); website: <http://www.uwstructuralgenomics.org>. This research funded by NIH/NIGMS Protein Structure Initiative grants USA GM074901 and P50 GM064598.

