Wheat Germ Cell-Free Eukaryotic Protein Production

Dmitriy A. Vinarov, Carrie L. Loushin Newman, Paul G. Blommel, Brian G. Fox, Jikui Song, and John L. Markley

University of Wisconsin-Madison Biochemistry, 433 Babcock Drive, Madison, Wisconsin, USA 53706-1549, http://www.uwstructuralgenomics.org

Abstract

CESG uses an automated platform [1] for wheat germ cell-free production of labeled proteins. Our current robotic systems (CellFree Sciences Co., Ltd., Japan) include the GeneDecoder1000™ (5-10 ug per well scale in 96-well format), the Protemist10™, and the Protemist100™ (1-2 mg per sample in eight samples format). Our cumulative experience with cell-free expression includes over 722 different structural genomics targets from human, mouse, *Plasmodium*, and *Arabidopsis*. Of 174 human targets that were successfully cloned, 135 (78%) showed expression levels suitable for structural investigations, 55 (41%) were soluble at concentrations needed for NMR, 36 (66%) gave purified [15N]-labeled samples at concentrations that could be evaluated by 1H-15N HSQC, and 10 of these gave favorable HSQC evaluations for foldedness and stability. Four of these human proteins have yielded NMR structures deposited in the PDB (five more are in progress). In total, CESG has deposited 12 eukaryotic proteins produced by wheat germ cell-free methodology (six more are in progress) into the PDB. The average yield of labeled purified proteins has been ~1.2 mg per mL of wheat germ lysate (OD_{260}=200). Preliminary studies show that the Protemist10™ can be used for automated preparation of sufficient quantities of Se-Met labeled protein for Fluidigm-based crystallization screening. We also report that cell-free translation provides a cost-effective and rapid method to screen multiple constructs engineered to improve solubility or foldedness. Using the Protemist10™ platform, expression and purification of 1-5 mg quantities of [^{15}N]-labeled protein for NMR structural studies is rapid and easy. Custom cell-free and cell-based expression vectors [2] utilizing the Flexi®Vector cloning system from Promega Corporation (Madison, WI) provide complementary methods for screening and production of target proteins with cleavable His6-, GST-, MBP-, and other tags in order to select the one most appropriate for production efforts. In cases where direct comparisons have been made with *E. coli* cell-based methods [3], the wheat germ system provides significant time and cost advantages for well-expressed and stable proteins, for [^{13}C, ^{15}N] labeling, and for advanced amino acid specific labeling strategies. The best pathway(s) for expression of more problematic targets remains to be elucidated.

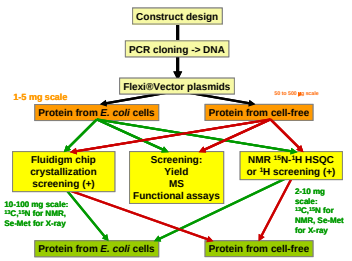
Center for Eukaryotic Structural Genomics (CESG) and Nuclear Magnetic Resonance Facility at Madison (NMRFAM)

Wheat Germ Cell-Free Protein Production Workshop

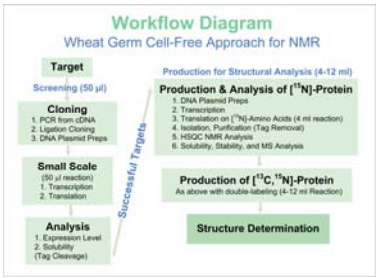
July 30 - August 4, 2006

- Hands-on experience with wheat germ cell-free
- Small-scale expression screening
- Large-scale production of labeled proteins
- Manual and automated procedures
- Try approaches with your own targets
- Only 24 participants and geared towards those w/3-5+ years of exp protein production/purification
- Registration deadline is: May 15, 2006
- $500.00 (academic) or $750.00 (private industry)

CESG's Cell-Based and Cell-Free Screening and Protein Production Pipelines for X-ray and NMR



Wheat Germ Cell-Free Approach for NMR



Automation: Robotic Systems from CellFree Sciences Co., Ltd. Japan

Gene Decoder™ Screening Mode	Protemist™ Large-Scale Protein Production Mode
4 x 96-well plates:	***8 x 4 ml reactions:***
• Overnight run	• Overnight run
• 2-5 µg protein / well	• 1 – 3 mg protein / reaction
• Uses 0.9 mL of wheat germ extract / plate	• Uses 3 ml of wheat germ extract / reaction

Expression Vectors

N-(His)$_6$ – FLEXI-pEU Vector - uncleavable

SP6 - Ω - (His)$_6$ - MCS - AA added to the N-terminus: MG(H)$_6$LE

N-(GST) – FLEXI- pEU Vector – cleavable with PreScission™ Protease

SP6 - Ω - GST - PreScission™ - MCS - AA added to the N-terminus: GPLE

N-(His)$_6$ -- pEU Vector – cleavable with PreScissionTM Protease

SP6 - Ω - (His)$_6$ - PreScissionTM - MCS - AA added to the N-terminus: GPLE

CESG's Efficient Automated Cell-Free Pipeline for NMR Targets

Small-scale (µg), automated 96-well format production overnight					Large-scale (mg), automated 8 X 4 mL production overnight				
Genome	Targets selected	Targets cloned successfully	Targets showing acceptable expression	Targets showing adequate solubility	[^{15}N]-labeled products produced	Acceptable ^{15}N-'H HSQC spectrum	Protein stable for >31 days	[^{13}C,^{15}N]-Labeled protein made‡	3D structures by NMR
Human	191	174 91%	135 70%	55 29%	36 19%	15 78%	10 5.2%	10 5.2%	9* 4.7%
Mouse	150	129 86%	47 31%	14 9.3%	11 7.3%	2 1.3 %	1 0.7%	1 0.7%	1 0.7%
Arabidopsis	381	351 92%	269 70%	120 31%	76 20%	17 4.4%	9 2.3%	9 2.4%	8** 2.1%
Total	722	654 91%	451 62%	189 26%	123 17%	34 4.7%	17 2.4%	13 1.8%	18 2.5%

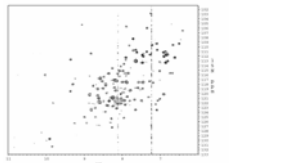
All percentages represent number of targets remaining relative to the number of targets selected.
‡ Average yield of purified double-labeled proteins used in structural investigations was 0.4 mg / mL.
* Includes five structures in progress. ** Includes two structures in progress.

E H – high expression (> 0.5 mg/ml); S H – high solubility (>75%)

Production and Prescreening of ^{15}N-Proteins

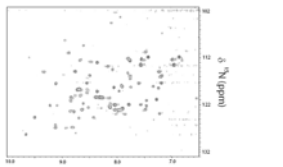
We are currently working on developing methodology for production of ^{15}N-proteins on a small-scale for prescreening foldedness. Two approaches are being investigated: (1) protein production on 'Protemist'™, utilizing only one 4-ml reaction, and (2) protein production on 'GeneDecoder 1000'™, utilizing 16 wells (16 x 150 µl reactions) per sample. Potentially, both approaches offer sizable savings since a significant number of unfolded, or otherwise unsuitable for NMR structure determination proteins, will be eliminated at this stage. Preliminary results from both approaches are shown below.

Spectrum of a ^{15}N-protein prepared on Protemist ™ utilizing only one 4-ml reaction:

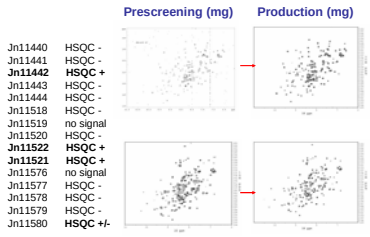


Hs.500165 (22 KDa) was synthesized (~ 500 µg) on 'Protemist'™ utilizing only one 4-ml reaction. The spectrum was collected on a 500 MHz Bruker spectrometer equipped with a cryogenic probe (16 transients with 128 complex points in the indirect [^{15}N] dimension; total acquisition time 1 h 20 min).

Spectrum of a ^{15}N-protein prepared in small-scale protein production mode on GeneDecoder™1000:



At2g24940.1 (13.5 KDa) was synthesized (~80-100 µg) in the small-scale protein production mode on the 'GeneDecoder 1000'™. Sixteen wells were devoted to this protein, with ~5 µg/well average yield. The spectrum was collected on a 500 MHz Bruker spectrometer equipped with a cryogenic probe (160 transients with 64 complex points in the indirect [^{15}N] dimension; total acquisition time 6 h).



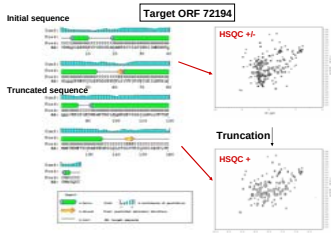
Truncations R&D

Full Length	Truncated					
E H S 0	E H S H	^{15}N Made	HSQC +	In Stability	Stability +	To Structure
42	5 (12%)	5 (12%)	0			
HSQC +/-	E H S H	^{15}N Made	HSQC +	In Stability	Stability +	To Structure
17	3 (18%)	3 (18%)	2 (12%)	2 (12%)	2 (12%)	2 (12%)
HSQC -	E H S H	^{15}N Made	HSQC +	In Stability	Stability +	To Structure
8	3 (38%)	3 (38%)	1 (13%)	1 (13%)	1 (13%)	1 (13%)
Total						
67	11 (16%)	11 (16%)	3 (4.5%)	3 (4.5%)	3 (4.5%)	3 (4.5%)

Truncation rescue of HSQC +/- targets: 2 / 17 (12%)

E H – high expression (> 0.5 mg/ml); S H – high solubility (>75%)

Modification of the target sequence provides an alternative method of rescuing marginal targets for NMR structure determination



Summary of the cleavable His-tag results

Main ID	MW	E	S	Scl	HSQC	Stability	^{13}C,^{15}N Protein	Comment
BC002752	13332	H	H		HSQC +/-			work stopped
BC009474	16703	H	H		HSQC +	stability +	+	in data collection
BC059148	13412	H	0					
BC030836	17565	H	0					
At4g16710.1	19582	H	0					
At5g03930.1	19869	L	H					
At3g28940.1	19464	H	H		HSQC +/-			work stopped
At2g14110.1	21704	L	H					
At3g09800.1	19738	L	L					
BC076248	19586	L	0					
Mm.307102	18065	H	L					
BC002949	20198	H	H		HSQC +	stability +	+	in data collection
BC006282	20747	0						
BC012478	20214	H	H		HSQC +/-			work stopped
BC007662	20274	H	H		HSQC +/-			multimer; Fluidigm
BC076368	13454	M	H		HSQC +/-			work stopped
At4g16710.1	18203	M	0					
BC005938	9937	M	M		HSQC +	stability -		protease inhibitor search
PF13_0098	16525	M	H		HSQC +/-			work stopped
PF14_0443	19310	M	H		HSQC +/-			work stopped

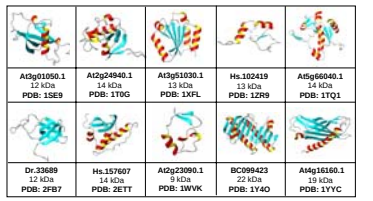
All of these targets with His-tag were HSQC +/

Cleavable His-tag rescue of HSQC +/- targets: 2 / 20 (10%)

Truncation vs. cleavable His-tag rescue results for original HSQC +/- targets

Main ID	Parent MW	Chamcik MW	E	S	Scl	HSQC	Cleavable His Tag	E	S	Scl	HSQC	^{13}C, ^{15}N
BC002752	13332	Δ347 8035	H	H		HSQC -		H	H	H	HSQC +/-	
BC009474	16703	Δ**N57**Δ**C119** 12251	M	0				H	H	H	HSQC +	delivered; in data collection
BC011683	19108	ΔN(Δ)C135 12929	H	0			cloning -					
BC059148	13412	ΔN(Δ)C43 7090	H	0				H	0			
BC030836	17565	ΔN11 16384	H	0				H	0			
At4g16710.1	19582	ΔN10 18322	H	0				H	0			
At5g03930.1	19869	ΔN52 14131	H	0				L	H	H		
At3g28940.1	19464	ΔN(Δ)C138 18287	H	M				H	H	H	HSQC +/-	
At2g14110.1	21704	ΔN(Δ)C27 13745	H	0				L	H	H		
At3g09800.1	19738	ΔN6ΔC44 14327	H	L				L	L	L		
BC076248	19586	ΔN81ΔC24 9783	H	0				L	0			

Wheat Germ Cell-Free Structure Gallery



Selected Publications

1. Vinarov, et al. (2005) High-throughput automated platform for NMR-based structural proteomics, *Expert Rev. Proteomics*, **2**, 49-55.
2. Blommel, et al. (2006) High efficiency single step production of expression plasmids from cDNA clones using the FlexiVector cloning system, *Prot. Express. Purif.*, in press.
3. Tyler, et al. (2005) Comparison of cell-based and cell-free protocols for producing target proteins from *Arabidopsis thaliana* for structural studies, *Proteins*, **59**, 633-643.

Acknowledgments

All CESG staff members and collaborators

CESG is supported by the National Institute of General Medical Sciences through the Protein Structure Initiative (1 U54 GM074901)