Center for Eukaryotic Structural Genomics

Protein Structure Initiative

High-Throughput Automated Platform for NMR-Based Structural Proteomics

Dmitriy A. Vinarov, Ejan M. Tyler, Mark N. Shahan, Carrie L. Loushin Newman, and John L. Markley

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

Abstract

The Center for Eukaryotic Structural Genomics (CESG) has developed an automated platform for NMR-based structural proteomics that employs wheat germ extracts for cell-free production of labeled proteins (1). The platform utilizes a single construct with uncleavable (N)-His_6 for all targets without any redesign of the DNA or RNA. The protein production and labeling protocol is no more costly than *E. coli* cell-based approaches for preparation of ^{15}N- and $^{13}C,^{15}N$-samples, is robust and scalable for high-throughput applications, and has been used in our Center to screen eukaryotic ORFs from the *Arabidopsis thaliana* (250), mouse (80), and human (160) genomes for expression and solubility and for the determination of NMR structures (four solved and deposited to PDB; six in progress). Given sufficient manpower and equipment for cloning and protein purification, CESG's 'GeneDecoder 1000™ and 'Protemist™ (Cell-Free Sciences, Co., Ltd.) robotic systems could carry out as many as 384 small-scale (125 µl each in the bi-layer mode) screening reactions per week and 24 production scale (4 ml each) reactions per week. At its current staffing level (four persons), CESG's cell free group can clone and screen 96 proteins every six weeks for protein production and solubility and deliver on average four ^{15}N-labeled proteins and one $^{15}N,^{13}C$ protein sample per week to the NMR spectroscopists. We will present our latest results and experience in utilizing this automated platform technology to screen targets for expression and solubility and to produce stable-isotope labeled samples for NMR structure determinations.

Rationale for exploring this technology

Structural proteomics projects require the expression and purification of thousands of proteins and/or protein fragments. The ability to obtain labeled proteins is an essential requirement for rapid progress in structure determinations. Critical examples are: (1) the production of U-^{15}N labeled protein to assess the foldedness and aggregation state; (2) the production of U-$^{13}C;^{15}N$, or selectively labeled proteins for NMR structure determinations; and (3) the production of Se-Met labeled protein for phase determination in X-ray studies. The successful implementation of cell-free protein expression may minimize problems in cell harvesting, cell lysis, and pre-column manipulations. Moreover, this approach may simplify purification, because the protein of interest is isolated from a smaller set of contaminants. Cell-free systems may also permit labeling strategies that cannot be achieved in whole cell systems, while potentially providing a substantial economy in the labeled material required to produce a target protein. This warrants serious consideration of how cell-free translation can contribute to high-throughput structural biology.

Advantages of the wheat germ cell-free protein production

- Supports rapid and efficient screening (supported by robotics)
- Requires smaller volumes (avoids lengthy concentration steps in protein purification)
- Labeled proteins can be prepared rapidly (in 1-2 days) to meet the needs of structural biologists
- Supports labeling strategies that are not practical for proteins produced from bacterial cells (no label scrambling)
- Supports the production of eukaryotic N-terminal $(His)_6$ proteins (previous experience showed that these were not produced successfully from *E. coli* cells)

Disadvantages of the wheat germ cell-free protein production

- Reagent intensive
- Currently not compatible with Gateway Cloning Technology used in the *E. coli* cell-based pipeline of the CESG project

CESG's wheat germ cell-free protein expression project involves two groups

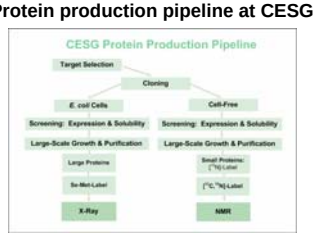
University of Wisconsin-Madison

- Screening of potential targets from eukaryotic genomes for suitability for structural studies
- Production of labeled proteins on the scale of several milligrams
- Assessment of this approach for high-throughput structure determination
- Improvement of the technology through its use in a production environment

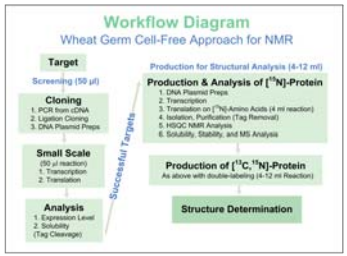
Ehime University and Cell-Free Sciences Co., Ltd.

- Wheat germ extract production
- Enabling methodology
- Robotics

Protein production pipeline at CESG



Wheat germ cell-free approach for NMR



Expression vectors

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

- SP6 - Ω - $(His)_6$ - MCS - AA added to the N-terminus: $MG(H)_6LE$

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

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

Expression and solubility screening of ORFs with N-terminal $(His)_6$ tag

Source	Expression +		Expression -		Solubility +		Solubility -	
Arabidopsis	263	70%	116	30%	119	45%	144	55%
Human	213	74%	75	26%	85	40%	128	60%
Mouse	66	46%	78	54%	18	27%	48	73%
Total	542	63%	269	37%	222	37%	320	63%

"Solubility +" indicates that target protein is at least 75% soluble.

Overall success rate for producing highly-soluble protein with N-terminal $(His)_6$ tag:

Arabidopsis: 30% Human: 30% Mouse: 12%

Expression and solubility screening of ORFs with N-terminal GST tag

Source	Expression +		Expression -		Solubility +		Solubility -	
Arabidopsis	86	79%	23	21%	56	65%	30	35%
Human	78	84%	15	16%	40	51%	38	49%

"Solubility +" indicates that fusion construct is at least 75% soluble.
For Human, 34 / 40 target proteins remained > 95% soluble after tag cleavage.
For *Arabidopsis*, 55 / 56 target proteins remained > 95% soluble after tag cleavage.

Overall success rate in expressing soluble protein by the cleavable GST route was:

Arabidopsis: 49% Human: 41%

Structural investigations

Source	Low Yield No HSQC		Higher Yield HSQC possible		Folded Non-Aggregating		Unsuitable for NMR structure		# Structures
Arabidopsis	51	50%	51	50%	23	45%	28	55%	6*
Human	10	30%	23	70%	9	40-%	14	60%	3**
Mouse	1	8%	11	92%	2	18%	9	82%	1

*One *Arabidopsis* structure is in progress; ** Three human structures are in progress.

Average yield for proteins used in structural investigations – 0.5 mg/ml

Comparison of target progress through three CESG pipeline protocols

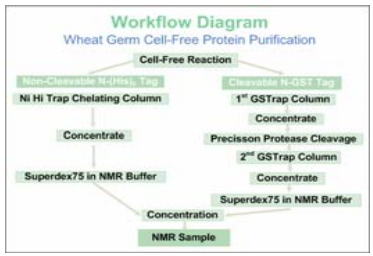
Ultimate step reached	Number of targets reaching a given ultimate step											
	E. coli MBP-tag			Cell-free His-tag			Cell-free GST-tag			Cell-free (best of both tags)		
	All	≤ 21	> 21	All	≤ 21	> 21	All	≤ 21	> 21	All	≤ 21	> 21
Sequence – (Seq)	0	0	0	0	0	0	0	0	0	0	0	0
Destination – (D)	1	1	0	0	0	0	0	0	0	0	0	0
E = Expression – (E)	39	19	20	14	6	8	12	10	2	1	5	2
Solubility – (S)	2	1	1	32	13	19	39	22	17	32	16	16
Cleavage – (C)	16	10	6	0	0	0	0	0	0	0	0	0
Purification – (P)	4	1	3	5	4	1	0	0	0	4	3	1
Low yield (L)	16	1	15	25	11	14	22	8	14	24	7	17
NMR sample prep – (N)	0	0	0	2	1	1	0	0	0	0	0	0
HSQC acquired	18	12	6	18	10	8	21	13	8	28	17	11
HSQC -	11	6	5	8	4	4	11	7	4	17	10	7
HSQC +/-	2	1	1	2	1	1	2	1	1	7	3	4
HSQC +	5	5	0	8	5	3	7	5	2	4	4	0

Ultimate step reached	Cell-free (best of both tags)			Best of all protocols E. coli and cell-free		
	All	≤ 21	> 21	All	≤ 21	> 21
Sequence – (Seq)	0	0	0	0	0	0
Destination – (D)	0	0	0	0	0	0
E = Expression – (E)	1	1	0	1	1	0
Solubility – (S)	32	16	16	32	16	16
Cleavage – (C)	0	0	0	7	6	1
Purification – (P)	4	3	1	6	2	4
Low yield (L)	24	7	17	27	8	19
NMR sample prep – (N)	0	0	0	1	0	1
HSQC acquired	28	17	11	35	20	15
HSQC -	17	10	7	24	13	11
HSQC +/-	3	3	0	1	1	0
HSQC +	8	4	4	10	6	4

Results are shown for all proteins and separated for the 45 proteins ≤ 21 kDa and the 51 proteins > 21 kDa.

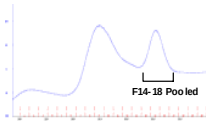
In general, the wheat germ cell-free platform offers the advantage of greater genome coverage for NMR-based structural proteomics, whereas the *E. coli* platform when successful yields more protein, as currently needed for crystallization trials for X-ray structure determination.

Wheat germ cell-free approach for NMR

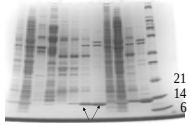


His-Tag purification of At2g24940 (14 kDa)

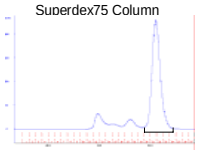
Ni-HiTrap Elution Profile



Ni-HiTrap Column Fractions



Superdex75 Column

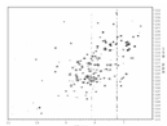


Superdex75 Fractions

Production and prescreening of ^{15}N-proteins

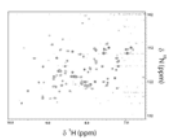
We are currently working on developing methodology for production of ^{15}N-proteins on small scale for prescreening foldedness. Two approaches are being investigated: (1) protein production on 'Protemist'™ utilizing only one 4-ml reaction and (2) protein production on 'GeneDecoder 1000'™ utilizing 16 wells (16 x 150 µl reactions) per sample. Potentially both approaches offer sizable savings since significant number of unfolded or otherwise unsuitable for NMR structure determination proteins will be eliminated on 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 (14 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).

Gene Decoder1000™

Screening Mode

4 x 96-well plates:

- Overnight run
- 2-5 µg protein / well
- Uses 2.5 – 5 mL of wheat germ extract / plate

Small-Scale Protein Production Mode

2 x 96-well plates:

- Overnight run
- 10-20 µg protein / well
- Uses 5 – 10 mL of wheat germ extract / plate

Protemist10™

Large-Scale Protein Production Mode

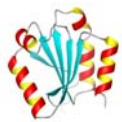
8 x 4 ml reactions:

- Overnight run
- 1 – 3 mg protein / reaction
- Uses 3 ml of wheat germ extract / reaction

Wheat germ cell-free structure gallery

At3g01050.1
12 kDa

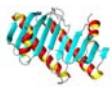
At2g24940.1
14 kDa



At3g51030.1
13 kDa

At5g66040.1
14 kDa

At2g23090.1
9 kDa



MGC37100
22 kDa

People involved

University of Wisconsin-Madison

Cell-Free Protein Production: John Markley, Dmitriy Vinarov, Ejan Tyler, Mark Shahan, Carrie Loushin Newman

NMR: Min Lee, Claudia Cornilescu, Rob Tyler, Jikui Song, Shanteri Singh, Qin Zhao

Medical College of Wisconsin, Milwaukee

NMR: Brian Volkman, Betsy Lytle, Francis Peterson

CESG is supported by the National Institute of General Medical Sciences through the Protein Structure Initiative (P50 GM 64598)