Center for Eukaryotic Structural Genomics

Protein Structure Initiative

A "CONTROL WORKGROUP" OF 24 PROTEIN TARGETS USED TO TEST PLATFORM CHANGES

David J. Aceti, Craig A. Bingman, Ronnie O. Frederick, Katarzyna A. Gromek, Karl W. Nichols, Shin-ichi Makino, Russell L. Wrobel, Frank C. Vojtik, John G. Primm, Brian G. Fox, George N. Phillips, Jr., and John L. Markley

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

Abstract

The Control Workgroup consists of a selected set of proteins whose behavior is known at all stages of our protein production and structure determination platform. It has profound utility in defining the efficacy of any new method developed, which includes downstream consequences on expression, stabilization, purification, functional assay, and structure determination. Control workgroups also provide an essential project resource for process validation, error checking, and staff training, and their existence will also facilitate the dissemination of newly developed expression and purification methods to the broader research community. Furthermore, by inclusion of proteins whose successful crystallization conditions or NMR HSQC resonance assignments are already known, the use of control workgroups in this project extends our validation of methods to include cross-correlation along the entire breadth of the structural biology pipeline from selection of targets to deposition of coordinates in the PDB. The current Control Workgroup is weighted towards soluble proteins that behave well in our hands, which makes it particularly useful for guarding against changes that degrade the performance of the platform (a "do no harm" philosophy). The proteins assembled come from a variety of source organisms and exhibit a wide range of physical characteristics such as molecular weight and pI. Many are easily detectable by color arising from the presence of metals or cofactors or by simple activity assays. The Control Workgroup has been used to test how new expression vectors and growth medium formulations affect total expression, solubility, susceptibility to processing by TEV protease, and yield of purification. All experimental efforts on the Control Workgroup are captured in the Sesame LIMS system. Here we illustrate the use of the Control Workgroup with a comparison of expression vectors that differ in cloning method, *lacI* repressor strength, and expression in *E. coli* versus a Wheat Germ Cell-Free system.

Results

Twenty-four targets were chosen to make up the Control Workgroup (Table 1). At the time when they were chosen, many had been characterized in terms of typical yields from expression and purification trials in Wheat-Germ Cell-Free or *E. coli* or both, crystals (red) or favorable HSQC data (blue) had been obtained (purple indicates both crystallization+ and HSQC+), and many structures had been solved by CESG (see PDB ID's).

Table 1. The Control Workgroup.

#	ORF ID	Protein Name	MW	Assay	Wheat Germ Cell-Free History	*E. coli* Cell-Based History
1	34382	cytoplasmic dynein light chain	10990		NMR structure **1Y4O**	Small amount purified
2	6042	unknown protein At1g77540.1	11736		None	NMR structure **2EVN**. X-ray structure **1XMT**
3	14751	thioredoxin h1	12673	Spec assay following NADPH/insulin-DS redox reaction	NMR structure **1XFL**	Small amount 15N purified
4	33810	zinc finger protein	13169		NMR structure **1ZR9**	None
5	81370	unknown heme-binding protein	16474	red-colored from heme binding	Expressed soluble	No crystals, HSQC-, good yield
6	2361	unknown protein At1g01470.1	16543		HSQC+	NMR structure **1XO8**, high yields, crystal
7	13193	unknown protein At3g03773.1	17366		None	Variable yields
8	11624	thioredoxin-like protein	17947		No expression	NMR structure **1X0Y**, high yields
9	91592	allene oxide cyclase variant1	19522		None	Did not cleave
10	91593	allene oxide cyclase variant2	21232		Expressed soluble	X-ray structure **1Z8K**
11	35683	cysteine dioxygenase 1	23026	HLPC assay for cysteine sulfinic acid formation	None	X-ray structure **2ATF**, moderate yield
12	605	phosphatase	24537	pNPP phosphatase assay	None	X-ray structure **1XRI**, good yields
13	91571	Enhanced C3 green fluorescent protein	26748	Green/Fluorescent	None	X-ray structure **2QU1**
14	91591	TEV protease	26922	Fluoresence anisotropy-based protease assay	None	High yields
15	74368	Pre-mRNA processing factor 24	27223	Gel mobility shift assay	None	X-ray structure **2GHP**, high yields
16	80048	sarcosine dimethylglycine methyl-xase	33324	Coupled spec assay following deamination of adenine	None	X-ray structure **2O57**, very high yield
17	37540	glyoxylate/hydroxypyruvate reductase	35668	NADPH-linked spectrophotometric assay	None	X-ray structure **2H1S**, low yield
18	79368	Aspartoacylase	35735	Couple spec. assay following deamination of aspartic acid	Expressed soluble	X-ray structure **2I3C**, low yield
19	70653	dimetal phosphatase	36645	pNPP Phosphatase Assay	None	X-ray structure **2NXF**, variable yields
20	7312	putative steroid sulfotransferase	37140		None	X-ray structure **1Q44**
21	8210	12-oxophytodienoate-10,11 reductase	42691	NADPH-dependent reduction of TNT	None	X-ray structure **1Q45**, high yield
22	24674	agmatine iminohydrolase	43156	Assay of ammonia product by Bertelot reaction	Expressed soluble	X-ray structure **1VKP**, very high yields
23	34351	unknown protein	50256		None	X-ray structure **2GNX**, low yield
24	74329	Photinus (Firefly) luciferase	60844	Luciferase assay	None	None

The Control Workgroup was cloned into each of five expression vectors to compare the vectors and the performance of the *E. coli* and Wheat Germ Cell-Free platforms. Major differences between vectors are outlined in **Table 2** *(*see also poster by Russell Wrobel *et al.)*. Most importantly, it was of interest to determine;

- The viability of moving to a single, unified Flexi-cloning system for the *E. coli* and Wheat Germ Cell-Free platforms, from parallel Gateway (recombinant) and restriction cloning systems. This would result in great savings in time and money and also give identical N-terminal target sequences for better comparability.
- Whether the wildtype *lacI* repressor of pVP68K is indeed more compatible with CESG's autoinduction system than is the $lacI^q$ of pVP16, as suggested by previous research [1].
- The viability of the pVP65K "self-cleaving'" vector, which constitutively expresses low-levels of TVMV protease for *in vivo* cleavage of the Maltose Binding Protein (MBP) solubility tag.
- Whether the TEV-cleavable His-tag of the Cell-Free vector pEU-His-FV is advantageous compared to the non-cleavable tag of pEU-His.

Table 2. Expression vectors and platforms tested with the Control Workgroup.

Vector	Cloning Method	Antibiotic	Promotor/ Repressor	Tag Cleavage	Fusion Protein
Base Vectors					
pVP16 (*E. coli*)	**Gateway**	**Amp**	**T5/lacI^q**	**TEV**	**His- MBP** -TEV- **S- ORF**
pEU-His (Cell-Free)	**Restriction**	**N/A**	**SP6**	**None**	**His** - **XX- ORF**
Experimental Vectors					
pVP68K (*E. coli*)	**Flexi**	**Kan**	**T5/lacI**	**TEV**	**His- MBP** -3CP-TEV-**S- ORF**
pVP65K (*E. coli*)	**Flexi**	**Kan**	**T5/lacI**	**Self (TVMV)/ TEV**	**MBP**-TVMV-**His**-TEV- **S- ORF**
pEU-His-FV (Cell-Free)	**Flexi**	**N/A**	**SP6**	**TEV**	**His** -TEV- **S- ORF**

The *E. coli* vectors were tested in small-scale screens and large-scale production using strain B834(pRARE2) and autoinduction, and proteins were labeled with 15N or selenomethionine (SeMet) and subjected to crystallization screens or HSQC analysis as indicated in **Figures 1**, **2**, and **3**.

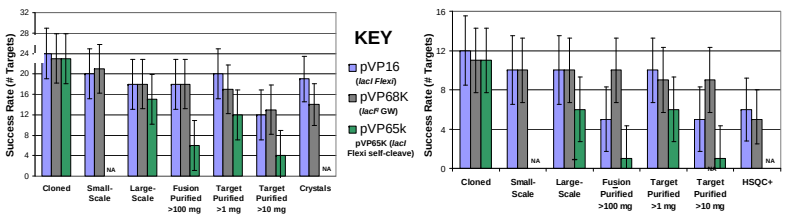

Figure 1. Performance of three *E. coli* vectors in the production of 24 SeMet-labeled Control Workgroup targets.

Figure 2. Performance of three *E. coli* vectors in the production of 12 ^{15}N-labeled Control Workgroup targets. Figure 2, above, suggested that fusion and target yields from the *lacI* pVP68K might be greater than those from the $lacI^q$ pVP16. To further study this effect, average yields of fusion and target are given in **Table 3** (in contrast with the number of targets above a threshold in **Figures 1** and **2**). The data in **Table 3** suggests that the pVP68K may indeed produce more protein than pVP65K, and that both produce more than pVP65K.

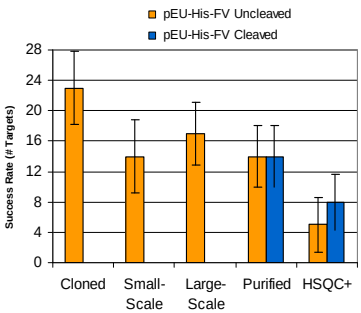

Figure 3. Performance of the Wheat Germ Cell-Free vector pEU-His-FV in the production of 24 ^{15}N-labeled Control Workgroup targets.

Table 3. Mean yield (mg) of fusion and target proteins from three *E. coli* vectors.

	^{15}N Fusion	SeMet Fusion	^{15}N Target	SeMet Target
pVP16 ($lacI^q$)	133	184	9	19
pVP68K (*lacI*)	281	255	30	26
pVP65K (*lacI* self-cleaving)	38	80	2	4

The wheat germ cell-free vector pEU-His-FV was tested with the 24 Control Workgroup targets (Figure 3). pEU-His-FV-expressed targets were carried forward through large-scale purification, then split into "Uncleaved" samples for immediate HSQC screening and "Cleaved" samples that were digested with TEV protease to remove the His tag, repurified, and subjected to HSQC analysis. Testing of the pEU-His vector is in progess and is not shown.

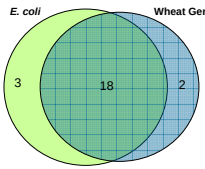

Figure. 4. Control Work Group Targets Expressed and Soluble from E. coli and Wheat Germ Cell-Free Systems. Are the *E. coli* and Wheat Germ Cell-Free systems overlapping or complementary with regard to the targets they are capable of expressing as soluble proteins? **Figure 4** addresses this question in the context of the newly formed the control workgroup. It should be noted that the control workgroup was created with a number of select characteristics but selection was not made to assure this result. This result reflects the findings of our earlier published work [2]; there are targets which are expressed and soluble in only one of our expression systems.

Conclusions

The Control Workgroup is a useful tool for comparing platform performance over a wide variety and a significant number of proteins. Using the Control Workgroup the following experimental results were facilely achieved:

- The lacI Flexi vector pVP68K "does no harm" and produces greater quantities of protein. The Flexi Cloning system works for both *E. coli* and wheat germ platforms. This system saves time and money.
- The "self-cleaving" vector pVP65K is not viable; the concept needs further research.
- The *E. coli* and Wheat Germ Cell-Free systems are complementary with regard to the targets they are capable of expressing as soluble protein.

** The Control Workgroup will be available shortly from the PSI Materials Repository as sets of 24 targets in four expression vectors.*

References

(1) Blommel PG, Becker KJ, Duvnjak P, and BG Fox. (2007) Enhanced bacterial expression during auto-induction obtained by alteration of Lac repressor dosage and medium composition. *Biotech Prog* 23, 585-598.

(2) Tyler RC, Aceti DJ, Bingman CA, Cornilescu CC, Fox BG, Frederick RO, Jeon WB, Lee MS, Newman CS, Peterson FC, Phillips GN, Jr, Shahan MN, Singh S, Song J, Sreenath HK, Tyler EM, Ulrich EL, Vinarov DA, Vojtik FC, Volkman BF, Wrobel RL, Zhao Q, Markley JL (2006) Comparison of cell-based and cell-free protocols for producing target proteins from *Arabidopsis thaliana* for structural studies. *Biochemistry* 45(48):14325-36

CESG is supported by the National Institute of General Medical Sciences through the Protein Structure Initiative NIH/NIGMS grant numbers U54 GM074901 (JLM; 07/01/05-06/30/10) and P50 GM064598 (JLM; 01/01/02-08/31/05).