Center for Eukaryotic Structural Genomics
Protein Structure Initiative

# The Center for Eukaryotic Structural Genomics: Technologies and Structures

J. Primm, D. Aceti, A. Bahrami, E. Beebe, L. Bergeman, S. Burgie, C. Bianchetti, C. Bingman, E. Bitto, R. Chylla, C. Cornilescu, B. Fox, R. Frederick, M. Goren, L. Grundhoefer, K. Gromek, D. Kondrashov, B. Lytle, S. Makino, J. Markley, Y. Matsubara , J. McCoy, K. Nichols, A. Nozawa, X. Pan, F. Peterson, G. Phillips, S. Sahu, D. Troestler, F. Vojtik, B. Volkman, G. Wesenberg, R. Wrobel, and Z. Zolnai

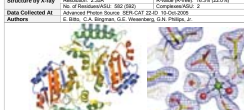
University of Wisconsin-Madison, Department of Biochemistry, 433 Babcock Drive, Madison, Wisconsin, USA **http://www.uwstructuralgenomics.org**

The Center for Eukaryotic Structural Genomics (CESG) aims to be a leading center for developing and disseminating tested technologies to solve structures of novel eukaryotic proteins. We create, evaluate, and optimize innovative protocols for producing eukaryotic proteins in active and soluble form. We refine methods for improving the yield of structures from high-value targets, in particular proteins from humans and other Vertebrates. We seek to improve the efficiency of all stages from target selection-design to three-dimensional structure determination by X-ray crystallography or NMR spectroscopy.

We actively share our advances with the biotechnology, pharmaceutical, and academic communities through collaborations, oral presentations, peer-reviewed articles, submissions to public databases and material repositories, and by conducting technology transfer workshops. Technologies (many of which will be presented) are tested by completing novel and community-requested structures for PDB, BMRB, and MR deposition. All CESG protein production protocols and Technology Dissemination reports are accessible at the PSI Knowledgebase and CESG's website.

## Center for Eukaryotic Structural Genomics
### Protein Structure Data Summary

| | | |
|---|---|---|
| **Target ID** | GO.36414 | |
| **Source Organism** | *Mus musculus* | |
| **Target Name** | BC038029 | |
| **PDB Entry** | 2BDU | Deposition: 20-Oct-2005 |
| **Function** | cytosolic 5′-nucleotidase III (FF/Refine: 2G06, 2G07, 2G08, 2G09, 2G0A, 2Q4T) | |
| **Produced From** | *E. coli* B834, pRARE2, pVP-16 | |
| **Structure by X-ray** | Resolution: 2.35Å | R-value (R-free): 16.3% (22.0%) |
| | No. of Residues/ASU: 582 (592) | Complexes/ASU: 2 |
| **Data Collected At** | Advanced Photon Source SER-CAT 22-ID 10-Oct-2005 | |
| **Authors** | E. Bitto, C.A. Bingman, G.E. Wesenberg, G.N. Phillips, Jr. | |



**Structural Features**

PN-1 catalyzes dephosphorylation of pyrimidine 5′-mononucleosides. Deficiency of PN-1 activity in erythrocytes results in non-spherocytic hemolytic anemia. This deficiency can be either hereditary, or acquired through lead poisoning. This is the first structure of a PN-1, and at 92% sequence identity to its human counterpart provides and excellent framework for understanding the human enzyme. The structure reveals a fold in the haloacid dehydrogenase superfamily. The active site is highly similar to phosphoserine phosphatases. In an extensive follow-up studies, we have defined the entire catalytic cycle of this enzyme through structures with $Mg^{++}$ alone, $Mg^{++}$ PO4, a beryllium fluoride phosphoenzyme analog, and an aluminum fluoride transition state analog. We have further structurally characterized the Lead(II) binding site, and propose a mechanism whereby Lead(II) compromises the active site, and inhibits this enzyme in acquired spherocytic anemia resulting from lead poisoning.

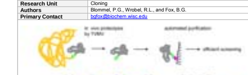
*References*: (1) Bitto, E., Bingman, C.A., Wesenberg, G.E., McCoy, J.G., Phillips, G.N., Jr. (2006) Structure of pyrimidine 5′-nucleotidase type 1. Insight into mechanism of action and inhibition during lead poisoning. *J Biol Chem* 281(29):20521-9.

| | |
|---|---|
| **Percent Identity with Nearest PDB Structure at Time Solved** | 16% (1NN1),10% (1QYI) |
| **Pfam Cluster** | UPMH-1 |
| **Sequence Cluster Size : Structures in PDB** | 70 NR at e<0.1 |

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: cesginfo@biochem.wisc.edu; website: http://www.uwstructuralgenomics.org. This research funded by NIH / NIGMS Protein Structure Initiative grants U54 GM074901 and P50 GM064598.

## Center for Eukaryotic Structural Genomics
### Technology Dissemination Report

| | |
|---|---|
| **CESG Tech Report No.** | 017 |
| **Title** | Customized Expression Vector Platform |
| **Research Unit** | Cloning |
| **Authors** | Blommel, P.G., Wrobel, R.L., and Fox, B.G. |
| **Primary Contact** | bgfox@biochem.wisc.edu |



**Summary**

A series of expression vectors used for protein production and different research activities are available [1-6]. These vectors use Gateway, FlexiVector, 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. The vector pVP65 is one example, and provides in vivo cleavage of MBP-target fusion proteins to liberate a His8-target that can be used with automated IMAC purification procedures. Sequence verified genes can be transferred between each of these different expression platforms by simple, highly efficient methods. These vectors are available by completion of standard biological materials transfer agreement, and will be deposited in the NIH PSI-Materials Repository.

Publication(s):

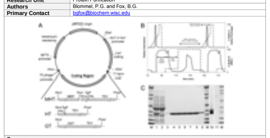
[1] Blommel, P.G., Martin, P.A., Wrobel, R.L., Steffen, E., Fox, B.G. (2006) High efficiency single step production of expression plasmids from cDNA clones using the Flexi Vector cloning system. *Protein Expr Purif* 47(2):562-70.
[2] Thao, S., Zhao, Q., Kimball, T., Steffen, E., Blommel, P. G., Riters, M., Newman, C. S., Fox, B. G., and Wrobel, R.L. (2004) Results from high-throughput DNA cloning of *Arabidopsis thaliana* target genes using site-specific recombination. *JSFG* 5(4):267-76.
[3] Blommel, P.G., Fox, B.G. (2005) Fluorescence anisotropy assay for proteolysis of specifically labeled fusion proteins. *Anal Biochem* 336(1):75-86.
[4] 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.
[5] Blommel, P.G., Martin, P. A., Seder, K. D., Wrobel, R.L., and Fox, B.G. (2007) Flexivector cloning. *Methods Mol Biol* 498:55-73.
[6] Frederick, R.O., Bergeman, L., Blommel, P.G., Bailey, L.J., Song, J., Meske, L., Bingman, C.A., Riters, 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., and Fox, B.G. (2007) Small-scale, semi-automated purification of eukaryotic proteins for structure determination. *JSFG* 8(4):153-66.

| | |
|---|---|
| **Acquiring the Technology** | Contact John Primm primm@nmrfam.wisc.edu. |
| **Other Acknowledgements** | Also supported by Promega Corporation, Madison WI (B. G. Fox, PI). |

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: cesginfo@biochem.wisc.edu; website: http://www.uwstructuralgenomics.org. This research funded by NIH / NIGMS Protein Structure Initiative grants U54 GM074901 and P50 GM064598.

## Center for Eukaryotic Structural Genomics
### Technology Dissemination Report

| | |
|---|---|
| **CESG Tech Report No.** | 015 |
| **Title** | Structural Genomics Methods Applied to Production of TEV Protease |
| **Research Unit** | Protein Purification |
| **Authors** | Blommel, P.G. and Fox, B.G. |
| **Primary Contact** | bgfox@biochem.wisc.edu |



**Summary**

Tobacco etch virus NIa 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 [1]. Small-scale screening, purification and mass spectral analysis showed that TEV protease with a C-terminal poly-Arg tag was proteolysed 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 fusion per g of *E. coli* cell paste) are also reported. These vectors are available by completion of standard biological materials transfer agreement, and will be deposited in the NIH PSI-Materials Repository.

Publication:

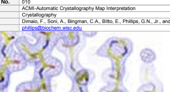
[1] Blommel, P.G. and Fox, B.G., A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expr Purif* 55(1):53-68.

| | |
|---|---|
| **Acquiring the Technology** | Contact John Primm primm@nmrfam.wisc.edu. |
| **Other Acknowledgements** | NIH Biotechnology Training Grant, T32 GM08349 to P.G. Blommel. |

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: cesginfo@biochem.wisc.edu; website: http://www.uwstructuralgenomics.org. This research funded by NIH / NIGMS Protein Structure Initiative grants U54 GM074901 and P50 GM064598.

## Center for Eukaryotic Structural Genomics
### Technology Dissemination Report

| | |
|---|---|
| **CESG Tech Report No.** | 010 |
| **Title** | ACMI-Automatic Crystallography Map Interpretation |
| **Research Unit** | Crystallography |
| **Authors** | Dimaio, F., Soni, A., Bingman, C.A., Bitto, E., Phillips, G.N., Jr., and Shavlik, J. |
| **Primary Contact** | phillips@biochem.wisc.edu |



**Summary**

One particularly time-consuming step in protein crystallography is interpreting the electron density map; that is, fitting a complete molecular model of the protein into a 3D image of the protein produced by the crystallographic process. In poor-quality electron density maps, the interpretation may require a significant amount of a crystallographer's time. We have investigated automating the time-consuming initial backbone trace in poor-quality density maps [1−3]. We describe ACMI (Automatic Crystallographic Map Interpreter), which uses a probabilistic model known as a Markov field to represent the protein. Residues of the protein are modeled as nodes in a graph, while edges model pairwise structural interactions. Modeling the protein in this manner allows the model to be flexible, considering an almost infinite number of possible conformations, while rejecting any that are physically impossible. Using an efficient algorithm for approximate inference — belief propagation — allows the most probable trace of the protein's backbone through the density map to be determined. We have tested ACMI on a set of density maps (at 2.5 to 4.0 Å resolution) and have shown that ACMI offers a more accurate backbone trace than current approaches.

Publication(s):

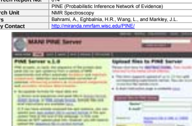
[1] DiMaio, F., Shavlik, J., and Phillips, G.N., Jr. (2003) Using pictorial structures to identify proteins I X-ray crystallography density maps. Working Notes of the ICML Workshop on Machine Learning, in *Bioinformatics*.
[2] DiMaio, F., Shavlik, J., Phillips, G.N., Jr. (2006) A probabilistic approach to backbone tracing in electron density maps. *Bioinformatics* 22(14):e81-9.
[3] DiMaio, F., Soni, A., Phillips, G.N., Jr. and Shavlik, J. (2007) Improved methods for template-matching in electron density maps using spherical harmonics. *IEEE-BIBM 2007 Conference Proceedings*: 258-265.

| | |
|---|---|
| **Acquiring the Technology** | ftp://ftp.cs.wisc.edu/machine-learning/shavlik-group/programs/acmi/ |
| **Other Acknowledgements** | Also supported by the National Library of Medicine LM007359 (PI:JS). |

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: cesginfo@biochem.wisc.edu; website: http://www.uwstructuralgenomics.org. This research funded by NIH / NIGMS Protein Structure Initiative grants U54 GM074901 and P50 GM064598.

## Center for Eukaryotic Structural Genomics
### Technology Dissemination Report

| | |
|---|---|
| **CESG Tech Report No.** | 014 |
| **Title** | PINE (Probabilistic Inference Network of Evidence) |
| **Research Unit** | NMR Spectroscopy |
| **Authors** | Bahrami, A., Eghbalnia, H.R., Wang, L., and Markley, J.L. |
| **Primary Contact** | http://miranda.nmrfam.wisc.edu/PINE/ |



**Summary**

PINE, which is available from a webserver at http://miranda.nmrfam.wisc.edu/PINE/, is the implementation of a multiple step probabilistic data analysis platform for protein NMR spectroscopy. As of 12/08 PINE has been used more than 1300 times by external Users. PINE incorporates the capabilities of separate probabilistic tools: PISTACIO [1] (automated backbone and sidechain assignment), PECAN[2] (secondary structure determination), and LACS [3] (referencing offset and outlier detection). The input to PINE is the amino acid sequence and sets of peak lists generated from one or more of the standard types of protein NMR experiments; these can be either probabilistic (e.g., peak lists generated by HIFI-NMR) or traditional peak lists generated by popular NMR data analysis tools. PINE takes into account the interconnectedness of different stages of analysis. PINE begins with a set of local statistical potentials. It then proceeds iteratively until a stationary state for a consistent global similarity measure is achieved. The resulting software enables a seamless and robust integration of multiple steps in the NMR structure determination pipeline. PINE provides as output a probabilistic assignment of backbone and sidechain signals and the secondary structure of the protein. At the same time, it identifies, verifies, and if needed rectifies, problems related to referencing, assignment, or outlying data. PINE can make use of prior information supplied from selective labeling or spin system assignments derived independently by other means. The performance of PINE is much superior to that of the individual tools used sequentially.

Publications:

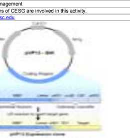
[1] Eghbalnia, H.R., Bahrami, A., Wang, L., Assadi, A., and Markley, J.L. (2005) *J Biomol NMR* 32(3):219-33.
[2] Eghbalnia, H.R., Wang, L., Bahrami, A., Assadi, A., and Markley, J.L. (2005) *J Biomol NMR* 32(1):71-81.
[3] Wang, L., Eghbalnia, H.R., Bahrami, A., and Markley, J.L. (2005) *J Biomol NMR* 32(1):13-22.

| | |
|---|---|
| **Acquiring the Technology** | Available from: http://miranda.nmrfam.wisc.edu/PINE/ |
| **Other Acknowledgements** | Partial support from NIH Grant P41 RR02301. |

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: cesginfo@biochem.wisc.edu; website: http://www.uwstructuralgenomics.org. This research funded by NIH / NIGMS Protein Structure Initiative grant U54 GM074901

## Center for Eukaryotic Structural Genomics
### Technology Dissemination Report

| | |
|---|---|
| **CESG Tech Report No.** | 018 |
| **Title** | Materials Distribution |
| **Research Unit** | Project Management |
| **Authors** | All members of CESG are involved in this activity. |
| **Primary Contact** | primm@wisc.edu |



**Summary**

CESG has worked with the PSI Materials Resource (PSI-MR) and the University of Wisconsin to develop acceptable material transfer agreements for this important aspect of sharing the results of the PSI. We have developed protocols for the physical transfer of plasmids containing CESG targets between our institutions. The effort included working with PSI-MR to define the format of the electronic documents to be transferred between the depositor and the PSI-MR prior to shipment of plasmids, establishing the string of check-offs that must occur before plasmids are shipped, and developing definitions and descriptors for experimental results associated with each plasmid deposition. The electronic process was deemed acceptable, and CESG then deposited 96 target clones into the PSI-MR as a pilot study for the physical transfer process. The physical transfer process was deemed acceptable. As of 7/1/2009 CESG has deposited more than 1500 plasmids have at PSI-MR which are available to the public. CESG is poised to deposit more than 4000 more in the fall of 2009.

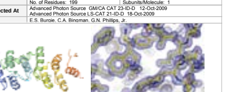
Our most requested plasmid is the optimized TEV protease expression vector described in CESG Technology Dissemination Report 015 (please see our website). As of October 2008, this was the single most requested material from PSI-MR. We also continue to have substantial distribution of our vectors pVP13 and pVP16 shown above. All CESG pipeline vectors are available through PSI-MR.

| | |
|---|---|
| **Acquiring the Technology** | Contact John Primm primm@wisc.edu or the PSI Materials Repository http://www.hip.harvard.edu/PSIMR/index.htm |
| **Other Acknowledgements** | |

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: cesginfo@biochem.wisc.edu; website: http://www.uwstructuralgenomics.org. This research funded by NIH / NIGMS Protein Structure Initiative grants U54 GM074901 and P50 GM064598.

## Center for Eukaryotic Structural Genomics
### Protein Structure Data Summary

| | | |
|---|---|---|
| **Target ID** | GO.88344 | |
| **Source Organism** | *Galdieria sulphuraria* | |
| **Target Name** | DCUN1 domain containing protein | |
| **PDB Entry** | 3KEV | Deposition: 26-Oct-2009 |
| **Function** | Presumably involved in protein neddylation | |
| **Produced From** | Wheat germ cell-free synthesis | |
| **Structure by X-ray** | Resolution: 1.84 Å | R-value (R-free): 15.6% (17.8%) |
| | No. of Residues: 199 | Subunits/Molecule: 1 |
| **Data Collected At** | Advanced Photon Source GM/CA CAT 23-ID-D 12-Oct-2009<br>Advanced Photon Source LS-CAT 21-ID-D 18-Oct-2009 | |
| **Authors** | E.S. Burgie, C.A. Bingman, G.N. Phillips, Jr. | |



**Structural Features**

This target was selected as part of a pair of workgroups, designed to compare outcomes from medically relevant human proteins, and homologs from the thermophilic eukaryote, Galdieria sulphuraria. In this case, the sequence of 3KEV is 50% identical to its human ortholog, whereas the identity between the closest structure in the PDB to the human ortholog was less than 30%. The human ortholog failed 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 mouse 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[1], Alzheimer's disease[2], and life-span[1]. It is also noteworthy that this was CESG's first non-control target protein to be synthesized by cell-free techniques that has progressed to the Protein Databank.

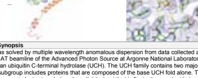
1. Gao, *et al.* (2006) Nature Cell Biology Oct;8(10):1171-7
2. Chen (2004) Apoptosis July;9(4):415-422

| | |
|---|---|
| **Percent Identity with Nearest PDB Structure at Time Solved** | 3BQ3 (29%IDE) |
| **Pfam Cluster** | DUF298 |
| **Protonet Cluster Size : Structures in PDB** | >250:2 |

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: cesginfo@biochem.wisc.edu; website: http://www.uwstructuralgenomics.org. This research funded by NIH / NIGMS Protein Structure Initiative grants U54 GM074901 and P50 GM064598.

## Center for Eukaryotic Structural Genomics
### Protein Structure Data Summary

| | | |
|---|---|---|
| **Target ID** | 41296 | |
| **Organism** | *Homo sapiens* | |
| **Gene Designator** | BC015521 | |
| **PDB Entry** | 3IHR | |
| **Function** | Ubiquitin hydolase | |
| **Produced From** | *E. coli* B834 P(RARE2) | |
| **Structure by X-ray** | Resolution: 50-2.95 Å | R-value (R-free): 20.1% (24.2%) |
| | No. of Residues/ASU: 328 | Subunits/ASU: 1 |
| **Data Collected At** | APS GM/CA-CAT 23-ID-D, 1 Dec 2008 | |
| **Authors** | Burgie, E.S., Bingman, C.A., Grundhoefer, L., Soni, A.B., Phillips Jr., G.N. | |



**Protein Synopsis**

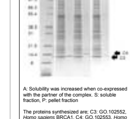
Uch37 was solved by multiple wavelength anomalous dispersion from data collected at the GM/CA CAT beamline of the Advanced Photon Source at Argonne National Laboratories. Uch37 is an ubiquitin C-terminal hydrolase (UCH). The UCH family contains two major divisions. The first subgroup includes proteins that are composed of the base UCH fold alone. This fold comprises a central, six-stranded, antiparallel β-sheet that is flanked on each side by α-helices. Enzymes from the second subgroup are composed of a UCH domain and a C-terminal extension. The C-terminal extensions may serve a variety of purposes, including, self-regulation, protein docking, and cellular localization. Uch37 is the first member of this subgroup to be characterized structurally. Its structure revealed that Uch37 forms dimers with C2 symmetry as the UCH domain of one subunit binds to the C-terminal extension of the other subunit. Furthermore, dimerization precludes the binding of ubiquitin to the UCH-domain, thus, explaining Uch37's auto-inhibitory mechanism. Uch37 provides an editing functionality for the 26S proteasome, which liberates poorly ubiquitinated proteins from destruction in the proteasome. Uch37 binds to SMAD7, which is a member of the transforming growth factor-β signaling cascade. In this arena Uch37 may protect TGF-β receptors from degradation. Uch37 also associates with nuclear factor related to κ-B of the Ino80 complex. Through this association Uch37 may provide a vital link between the proteasome and the ubiquitination state of Nucleosomes.

| | |
|---|---|
| **Percent Identity with Nearest PDB Structure at Time Solved** | 27% (1UCH) |
| **Pfam Cluster** | Peptidase_C12 |
| **Sequence Family Size** | 383 |
| **CESG Publications** | To be published |

Reference: Structural characterization of Uch37 reveals the basis of its auto-inhibitory mechanism. *To be Published*

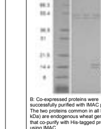
## Center for Eukaryotic Structural Genomics
### Technology Dissemination Report

| | |
|---|---|
| **CESG Tech Report No.** | 021 |
| **Title** | Co-expression of Protein:Protein complexes using a wheat germ cell-free system |
| **Research Unit** | Cell-Free Protein Production |
| **Authors** | Makino S., Primm J.G. and Fox B.G. |
| **Primary Contact** | jgprimm@wisc.edu |



A: Solubility was increased when co-expressed with the partner of the complex. S: soluble fraction, P: pellet fraction

The proteins synthesized are; C3: GO.102552, *Homo sapiens* BRCA1, C4: GO.102553, *Homo sapiens* BARD1



B: Co-expressed proteins were successfully purified with IMAC purification. The two proteins common in all lanes (~50 kDa) are endogenous wheat germ proteins that co-purify with His-tagged proteins using IMAC.

The proteins synthesized are;
C3: GO.102552, *Homo sapiens* BRCA1,
C4: GO.102553, *Homo sapiens* BARD1.

**Summary**

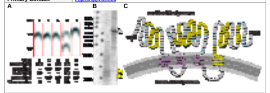
We have begun work on targets which form protein:protein complexes. To establish viability of our approach, a workgroup (WG) of known heterodimeric complexes was selected. This WG is being tested in both cell-based and cell-free expression systems. Using the test protein pair Homo sapiens BRCA1:BARD1, (PDB:1JM7) previously expressed in *E. coli*, we have seen marked improvement in solubility and ease of purification when the proteins are co-expressed in the cell-free system. In this experiment separate vectors were combined in a single cell-free reaction vessel which allowed the complex to form at the time of synthesis. When the targets are expressed without the partner, they are insoluble in the system.

| | |
|---|---|
| **Acquiring the Technology** | Contact Brian Fox bgfox@biochem.wisc.edu. |
| **Other Acknowledgements** | Not Applicable |

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: cesginfo@biochem.wisc.edu; website: http://www.uwstructuralgenomics.org. This research funded by NIH / NIGMS Protein Structure Initiative grants U54 GM074901 and P50 GM064598.

## Center for Eukaryotic Structural Genomics
### Technology Dissemination Report

| | |
|---|---|
| **CESG Tech Report No.** | 020 |
| **Title** | Activity and Topology Determination of *In vitro* Expressed Membrane Proteins |
| **Research Unit** | Cell-Free Protein Product |
| **Authors** | Goren M.A., and Fox B.G. |
| **Primary Contact** | mgoren@wisc.edu |



**Summary**

We report reconstitution of a three-component membrane protein complex using wheat germ cell-free technology [1]. Proteoliposomes containing the peripheral membrane protein cytochrome $b_5$ (CytB5) and the integral membrane protein human stearoyl-CoA desaturase isoform 1 (hSCD1) were produced in a cell-free expression reaction supplemented with 100 nm unilamelar liposomes derived from soy bean total lipid extracts. Purification of hSCD1/CytB5 proteoliposomes was accomplished by single step gradient floatation. Prior to initiation of stearoyl-CoA desaturation by addition of CytB5 reductase, NADH and stearoyl-CoA, the hSCD1 diiron and CytB5 heme cofactors were reconstituted by incubation with $Fe^{2+}$ and hemin chloride. Conversion of stearoyl-CoA (18:0) to oleoyl-CoA (18:1) was monitored by the addition of a [U-$^{14}$C]-stearoyl-CoA tracer, resolved by thin-layer chromatography, and visualized on a phospho-imager (Figure A).

Investigation of the topology of hSCD1 proteoliposomes produced *in vitro* is done in collaboration with Professor Yaeta Endo (Ehime University, Matsuyama, Japan) and Professor Lloyd Smith (University of Wisconsin-Madison). Initial identification of hSCD1 topology was performed by protease protection of [$^{14}$C]-leucine labeled protein. Floated proteoliposomes with (Fig. B, lane 3, yellow stars) and without (lane 4, white star) proteinase K pretreatment were visualized by autoradiography. Two fragments were protected from proteolysis, in agreement with the predicted topology containing ~5.5 and 6.5 kDa internal fragments. A more detailed analysis of topology was accomplished using an HPLC-ESI-MS/MS. Purified hSCD1 proteoliposomes were digested by trypsin during an 18-h time-course and loaded directly for analysis. As visualized in Figure C, 8 diagnostic fragments (yellow spheres) were observed, bracketed by trypsin cut sites (cyan spheres). Seven of these fragments were cytoplasmic peptides, while one, which was detected only at the final time point, was a predicted transmembrane domain. These data strongly support the utility of wheat germ cell-free technology to produce active and correctly folded integral membrane proteins.

Publication:

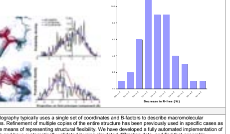
[1] Goren M.A. and Fox B.G. Wheat germ cell-free translation, purification, and assembly of a functional human stearoyl-CoA desaturase complex. *Prot Exp Purif* 62(2):171-8.

| | |
|---|---|
| **Acquiring the Technology** | Contact Brian Fox bgfox@biochem.wisc.edu. |
| **Other Acknowledgements** | Not Applicable |

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: cesginfo@biochem.wisc.edu; website: http://www.uwstructuralgenomics.org. This research funded by NIH / NIGMS Protein Structure Initiative grants U54 GM074901 and P50 GM064598. MAG was also supported by an NSF EAPSI post-doctoral international research fellowship.

## Center for Eukaryotic Structural Genomics
### Technology Dissemination Report

| | |
|---|---|
| **CESG Tech Report No.** | 009 |
| **Title** | Ensemble Refinements of Protein Structures |
| **Research Unit** | Crystallography |
| **Authors** | Levin. E.J., Kondrashov, D.A., Wesenberg, G.E., and Phillips, G.N., Jr. |
| **Primary Contact** | phillips@biochem.wisc.edu |




**Summary**

X-ray crystallography typically uses a single set of coordinates and B-factors to describe macromolecular conformations. Refinement of multiple copies of the entire structure has been previously used in specific cases as an alternative means of representing structural flexibility. We have developed a fully automated implementation of this approach and have systematically validated it using simulated diffraction data, and find that ensemble refinement produces better representations of the distributions of atomic positions in the simulated structures than single conformer refinements, and can accurately model anharmonic and multi-modal motions (left figure).[1] Comparison of principal components calculated from the refined ensembles and simulations shows that concerted motions are captured locally, but correlations dissipate over long distances. The technique was also tested on 50 experimental structures of varying resolution, and led to decreases in R-free (right figure), implying that improvements in the representation of flexibility observed for the simulated structures may apply to real structures. These gains were essentially independent of resolution or data-to-parameter ratio, suggesting even structures at moderate resolution can benefit from ensemble refinement.

Publication:

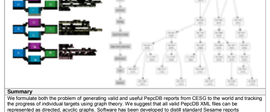
[1] Levin, E.J., Kondrashov, D.A., Wesenberg, G.E., and Phillips, G.N., Jr. (2007) Ensemble refinement of protein crystal structures: validation and application. *Structure* 15(9):1040-52.

| | |
|---|---|
| **Acquiring the Technology** | Send request by email to ejlevin@biochem.wisc.edu. |
| **Other Acknowledgements** | |

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: cesginfo@biochem.wisc.edu; website: http://www.uwstructuralgenomics.org. This research funded by NIH / NIGMS Protein Structure Initiative grants U54 GM074901 and P50 GM064598.

## Center for Eukaryotic Structural Genomics
### Technology Dissemination Report

| | |
|---|---|
| **CESG Tech Report No.** | 002 |
| **Title** | Target Trace and PepcDB Generation Software |
| **Research Unit** | Bioinformatics |
| **Authors** | Pan, X., Wesenberg, G.E., and Bingman, C.A. |
| **Primary Contact** | cbingman@biochem.wisc.edu |



**Summary**

We formulate both the problem of generating valid and useful PepcDB reports from CESG to the world and tracking the progress of individual targets using graph theory. We suggest that all valid PepcDB XML files can be represented as directed, acyclic graphs. Software has been developed to distill standard Sesame reports containing hierarchical (parent-child) relationships between database objects such as workgroups and sample records into visually appealing and intuitive pictures. These graphs simultaneously display all work done on a given target as one or more graphs rooted in target selection actions, and flowing across our multi-vector, multipass standard platform. There is no need for fixed templates to describe workflow. This allows our platform to be assembled literally "on the fly" based on the response of a given target protein to a battery of experimental trials, while maintaining the ability to report detailed PepcDB trials for each distinct experimental variation. The graph also allows users to easily pinpoint data entry errors, such as mislinked database objects. It has been invaluable in increasing the scientific accuracy and utility of our PepcDB reports.

Although this software was developed in the context of the Protein Structure Initiative, it is trivially extensible to describe the flow of functional proteomics work, or any other high-throughput activity using the Sesame database. With some additional effort, the software could be adapted to produce graphical reports from other project databases. The software was written using freely distributed components, such as Perl, DOT, and GraphViz software, and displays results via an Apache web server to a user's web browser.

Publication:

[1] Pan, X., Wesenberg, G.E., Markley, J.L., Fox, B.G., Phillips, G.N., Jr., and Bingman, C.A.. A graphical approach to tracking and reporting target status in structural genomics. *JSFG* 8(4):209-16.

| | |
|---|---|
| **Acquiring the Technology** | Contact Craig Bingman cbingman@biochem.wisc.edu. |
| **Other Acknowledgements** | The Sesame programming team, especially Zsolt Zolnai and all CESG Sesame users. |

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: cesginfo@biochem.wisc.edu; website: http://www.uwstructuralgenomics.org. This research funded by NIH / NIGMS Protein Structure Initiative grants U54 GM074901 and P50 GM064598.

## Center for Eukaryotic Structural Genomics
### Technology Dissemination Report

| | |
|---|---|
| **CESG Tech Report No.** | 022 |
| **Title** | Outreach Activities at CESG: Community Requests, CESG Web Portal and Publications |
| **Research Unit** | CESG |
| **Authors** | All members of CESG are involved in these activities |
| **Primary Contact** | jgprimm@wisc.edu |

**Summary**

**CESG Web Portal:** CESG began tracking hits on CESG web pages in September of 2008. Currently the site averages more than 1000 unique hits each month. Since tracking began, the site has been accessed from 1400 different cities from 100 different countries. http://www.uwstructuralgenomics.org.

**Community Requests:** During years 1-4 of PSI-2, 369 targets were proposed by outside requestors. Of these, 291 were accepted by CESG, 70 were rejected, and 8 are pending a decision. Of those accepted, 9 were full-length except for the removal of signal peptides, 67 were predicted domains, and the remaining 215 were full-length. Of those accepted 34 are still in play in the CESG expression platform. Eight structures have been derived from the outside request pool. CESG actively seeks structure nominations from the biology community. To submit a request please visit http://www.uwstructuralgenomics.org/structurerequest.htm

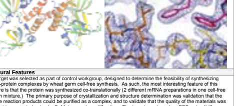
**Publications:** As of July 15th, 2009, CESG has published more than 70 method development articles 70 peer reviewed articles CESG solved structures. All CESG publications can be found on our website http://www.uwstructuralgenomics.org/publications.htm and through the PSI Knowledge Base http://kb.psi-structuralgenomics.org

| | |
|---|---|
| **Acquiring the Technology** | Contact John Primm jgprimm@wisc.edu |
| **Other Acknowledgements** | Not Applicable |

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: cesginfo@biochem.wisc.edu; website: http://www.uwstructuralgenomics.org. This research funded by NIH / NIGMS Protein Structure Initiative grant P50 GM064598.

## Center for Eukaryotic Structural Genomics
### Protein Structure Data Summary

| | | |
|---|---|---|
| **Target ID** | GO.102544, GO.102545 | |
| **Source Organism** | *Homo sapiens* | |
| **Target Name** | Human replication protein A complex | |
| **PDB Entry** | 3KDF | Deposition: 22-Oct-2009 |
| **Function** | Binds single-stranded DNA during replication | |
| **Produced From** | Wheat germ cell-free synthesis | |
| **Structure by X-ray** | Resolution: 1.98 Å | R-value (R-free): 19.2% (20.6%) |
| | No. of Residues: 506/ASU | Subunits/Molecule: 1 heterotetramer/ASU |
| **Data Collected At** | Advanced Photon Source GM/CA CAT 23-ID-D 05-Aug-2009 | |
| **Authors** | E.S. Burgie, C.A. Bingman, G.N. Phillips, Jr. | |



**Structural Features**

This target was selected as part of control workgroup, designed to determine the feasibility of synthesizing protein-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 (2 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 1LI0.

| | |
|---|---|
| **Percent Identity with Nearest PDB Structure at Time Solved** | 1LI0 (100%IDE) |
| **Pfam Cluster** | PF08784, PF08661 |
| **Cluster Size : Structures in PDB** | 164:3, 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: cesginfo@biochem.wisc.edu; website: http://www.uwstructuralgenomics.org. This research funded by NIH / NIGMS Protein Structure Initiative grants U54 GM074901 and P50 GM064598.

CESG is supported by the National Institute of General Medical Sciences through the Protein Structure Initiative NIH/NIGMS grant numbers 1 U54 GM074901 and P50 GM064598.