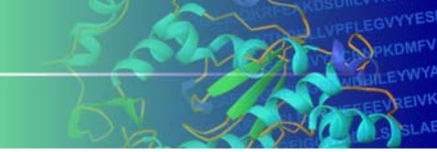


High-Throughput Crystallography at the Center for Eukaryotic Structural Genomics

Craig A. Bingman, Eduard Bitto, David J. Aceti, Lousie Meske, Jason McCoy,
Euiyong Bae, Zhaohui Sun, Gary Wesenberg, and George N. Phillips, Jr.
University of Wisconsin-Madison, Department of Biochemistry, 433 Babcock Drive, Madison, Wisconsin, USA 53706-1549, http://www.uwstructuralgenomics.org

Abstract

The Center for Eukaryotic Structural Genomics (CESG) has as its mission the dramatic reduction of costs and the lessening of the time required to determine a three-dimensional protein structure from eukaryotic organisms. We have developed and operate a high-throughput protein production pipeline using both *E. coli* and cell-free expression systems. Proteins for study are chosen based on general sequence dissimilarity to known structures, medical relevance, or are suggested by the larger scientific community. Although our initial set of proteins was chosen from the plant *Arabidopsis thaliana*, we now consider targets from any eukaryote. Approximately fifty structures have now been solved by X-ray crystallography including proteins not only from *Arabidopsis*, but also rice, mouse, zebrafish, and human origins. For X-ray crystallography, integration of robotic methods of screening, including microfluidic chip technology, sitting drop setups, and plate handling and imaging have led to an efficient crystal discovery system. Structure solution includes automatic, script-based cluster computing to produce optimal initial electron density maps. An overview of CESG crystallographic activities and connections between structure and function will be presented for selected X-ray structures. The work of the entire CESG staff is acknowledged.

See also related poster #112 on overall activities of CESG.

High-Throughput Crystallomics Platform

Screening and Optimization

All aspects of the screening and optimization process are supported by the Sesame LIMS system. Crystals are screened against a fixed composition 192 condition endogenous "general screen" named UW-192. Several commercially available screens are also used for salvage purposes. Worklists for the Tecan Genesis™ encoding the aspirate and dispense commands necessary to construct arbitrarily complex optimization plates are generated by the Well view of Sesame.

Crystal Imaging and Scoring

All imaging needs for standard pipeline screening are met by a pair of CrystalFarm™ imaging hotels, operating at 4°C and 20°C. These fully automated systems replace the semi-automated CrystalScore™ systems used in PSI Phase 1. Each has a total capacity of 400 crystallization plates, acquires images on a user-defined schedule, stores them on a local MySQL database running on a control computer with 0.5TB local storage, and presents them via a web-based interface for scoring.

New Sesame Features Supporting Crystallomics

Sesame has been the laboratory linformation management system for CESG since 2001. It collects, stores, manages, and mines data for all facets of our structural genomics proiject, from target selection to structure deposition. Standard output from Sesame forms the basis for our TargetDB and PepcDB depositions. Sesame currently has over 60 registered centers and labs around the world, including Promega Corporation, NMR Core Facility, Center for Biotechnology and Interdiscipliinary Studies, Rensselaer Polytechnic Institute, University of Melbourne, Australia, and the New York Consortium on Membrane Protein Structure (NYCOMPS).

In addition to the extraordinarily flexible, extensible and precise descriptions of crystallization experiments in the Well view, several new features have been recently added to facilitate emerging crystallomics strategies. Well now supports the volumetric merging of multiple screens, which enables high-throughput "perturbation" screens to be precisely defined in our database. This facility will enable routine deployment of these salvage screens in Phase 2.

Another new feature is the crystal view, shown above. This view provides direct links between crystal samples, crystallization conditions, and upstream details from protein purification, cloning and target selection. Well code has been extended to this view to enable both the declaration of cryoprotection solutions, and formulation of cryopotectants using the Tecan Genesis™.

• **go.74073 Dr.36843**

• Action trace	Date	User	Volume
• Selected	2005-03-25	csnewman	
• Primers Ordered	2005-03-29	csnewman	
• PCR +	2005-04-12	zteggers	
• Cloning Acceptable	2005-04-27	zteggers	
• Screening Acceptable	2005-06-07	johnk	
• ESC HHH	2005-06-23	koryseder	
• Paste Aacceptable	2005-06-30	fcvojtik	
• Purification Complete	2005-07-26	simond	4860
• QA Data Entered	2005-07-28	gsabat	4800
• QA Activated	2005-08-15	daceti	
• Screen Set Up	2005-08-24	zteggers	4350
• Crystallized +	2005-09-01	zteggers	
• Diffraction Crystals	2005-10-01	ebitto	4260
• Phasing Data +	2005-10-10	ebitto	4080
• X-ray PDB Deposited	2005-10-21	wesenberg	-780

Overall Project Statistics

Target Status	*Phase-1 End*	*2006-03-30*
Selected	6804	7416
PCR+	5454	5948
Sequence+	4709	5114
Screening Expression	2827	3151
Screening Soluble	2426	2669
Successfully Purified	578	714
To Crystallization	368	471
Crystal Structure/PDB	39	51
HSQC+	65	81
NMR Structure/PDB/BMRB	19	26

Fluidigm Topaz Microfluidics Crystallization

Motivations and Experimental Design

A forward-looking objective of our project is microscale protein expression and purification trials to prove targets prior to investing in large-scale growths and protein purification. One hundred microgram quantities of protein should become available for the incremental cost of concentration from the cell-based expression pathway. The cost of sample in cell-free expression already scales nearly linearly with target yield. The small quantities of protein from preliminary expression trials will be sufficient to run several hundred crystallization experiments in the Topaz™ free-interface diffusion environment, which requires around 10-14 nL of sample per experiment.

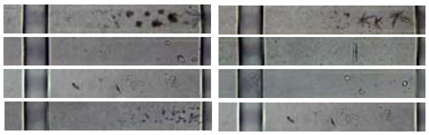
We have screened 16 unique targets and performed 88 distinct experiments. The primary objective has been to examine the suitability of Fluidigm as a "screen for crystallizability" by comparing results from free interface diffusion and vapor diffusion. We have also explored adapting our general screen to Fluidigm technology, by balancing water activity across individual chips. We are also evaluating the comparative efficacy of our general screen and Fluidigm reagents, and validating our micro-scale protein production process.

Preliminary Results

Based on the limited number of samples examined to date, a threshold result of three weak hits or one strong hit to trigger a large-scale growth and more extensive large-scale crystallization trials is appropriate. Under those conditions, Fluidigm screening would have produced a "go" signal for all targets crystallizable by vapor diffusion, and would have generated no false starts. Additional experiments are scheduled to more exhaustively define this threshold. Rigorous elimination of false-positives is important for optimizing a two-tier screening strategy. Our preliminary results show that the vast majority of positive results are apparent within 24 hours, and that continuing imaging beyond 3-4 days generates obvious false positives. Our preliminary results show no compelling advantage to using our 192 condition general screen or Fluidigm's proprietary reagents, after our screen was reformatted to control water activity.

Future Directions

Microfluidics crystallization screening is presently used for three purposes: screening samples produced by cell-free protein production, screening samples with costly, custom-synthesized components (e.g. nucleic acids) and screening low yield samples from cell-based protein production. In all three cases, sample or opportunity costs make microfluidics crystallization screening attractive.



Action Traces

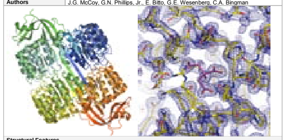
Standard reports are constructed from data in Sesame that describe the complete history of targets, from target selection to PDB deposition. The quantity of sample required to complete a structure is also tracked.

• **go.36731 Hs.104846 SSAT2**

• Action Trace	Date	User	Volume
• Selected	2005-03-08	csnewman	
• Sequence -	2005-04-18	ejan19	
• Primers Ordered	2005-04-29	zteggers	
• PCR +	2005-04-29	zteggers	
• Cloning Acceptable	2005-05-12	zteggers	
• Screening Acceptable	2005-06-22	johnk	
• ESC HHH	2005-07-11	koryseder	
• Paste Acceptable	2005-07-19	fcvojtik	
• Purification Complete	2005-08-10	breanne	2370
• QA Data Entered	2005-08-05	gsabat	
• QA Activated	2005-08-15	daceti	2310
• Screen Set Up	2005-08-23	zteggers	1860
• Crystallized +	2005-08-30	zteggers	
• Diffraction Crystals	2005-10-04	bwhan	1680
• Phasing Data +	2005-10-17	bwhan	1560
• X-ray PDB Deposited	2005-10-24	wesenberg	-810

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	78623 and 74631	
Organism	*Homo sapiens*	
Gene Designator	RNase/hRI Complex	
PDB Entry	1Z7X	Deposition: 21-Jun-2005
Function	pancreatic ribonuclease A complexed with inhibitor (EC 3.1.27.5)	
Produced From	*E. coli* BL21(DE3) pET-22B	
Structure by X-ray	Resolution: 1.95 Å	R-value (R-free): 17.8% (23.6%)
	No. of Residues/ASU: 1175	Complexes/ASU: 2
Data Collected At	Advanced Photon Source SBC 22-ID 12-Mar-2005	
Authors	J.G. McCoy, G.N. Phillips, Jr., E. Bitto, G.E. Wesenberg, C.A. Bingman	



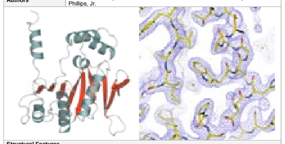
Structural Features

The x-ray crystal structure of human ribonuclease inhibitor (hRI) in complex with human pancreatic ribonuclease (RNase 1), solved in collaboration with the Ronald Raines lab (also at UW-Madison Biochemistry), gives a molecular picture of how the cell regulates the potentially damaging activity of RNase 1. It is the third structure of hRI in complex with a member of the ribonuclease A (RNase A) superfamily. By comparison to the other three structures, the hRI/RNase 1 complex shows the plasticity and generality of the leucine-rich repeat structure of hRI for accommodating, but still tightly inhibiting different members of the RNase A superfamily. Ribonucleases show promise as possible cancer therapeutics. Ribonucleases can be transformed into cancer therapeutics by blocking the inhibitory interaction of hRI. Creating variants of RNase 1 that evade hRI has been difficult, but this crystal structure may provide the important link to create cancer therapeutics using RNase 1.

Percent Identity with Nearest PDB Structure at Time Solved	91% over 127 aa (1E21) 95% over 460 aa (2BEX)
Pfam Cluster	RNase_A,
Sequence Cluster Size : Structures in PDB	449 and 883

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	36704	
Organism	*Homo sapiens*	
Gene Designator	Hs.433573	
PDB Entry	1ZTP	Deposition: 27-May-2005
Function	basophilic leukemia-expressed protein (BLES03, p5326)	
Produced From	*E. coli* B834, pRARE2, pVP-16	
Structure by X-ray	Resolution: 2.50Å	R-value (R-free): 18.8% (24.5%)
	No. of Residues/ASU: 686 (753)	Complexes/ASU: 3
Data Collected At	Advanced Photon Source NSLS X29A 05-May-2005	
Authors	E. Bitto, C.A. Bingman, H. Robinson, S.T.M. Allard, G.E. Wesenberg, G.N. Phillips, Jr.	



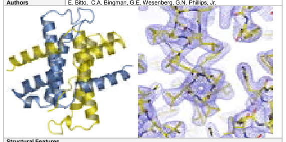
Structural Features

The crystal structure of BLES03 was solved by SeMet SAD using data collected at the mail-in program at NSLS. BLES03 shows no reliable sequence similarity to any functionally characterized proteins.
The structure of BLES03 adopts a fold similar to the eukaryotic transcription initiation factor 4E (eIF4E) a protein that binds the mRNA cap structure. In addition to fold similarity, the general features of the electrostatic surface potential of BLES03 and eIF4E are similar. It is proposed that BLES03 is involved in a biochemical process that requires recognition of nucleic acids. *Reference:* The structure at 2.5Å resolution of human basophilic leukemia-expressed protein BLES03. *Acta Cryst.* (2005). **F61**, 812-817.

Percent Identity with Nearest PDB Structure at Time Solved	9% (1IPB)
Pfam Cluster	Pfam-B_53833
Sequence Cluster Size : Structures in PDB	18 NR at e<0.1

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	34455	
Organism	*Mus musculus*	
Gene Designator	Mm.29898	
PDB Entry	2A3Q	Deposition: 25-Jun-2005
Function	probable hydrolase	
Produced From	*E. coli* B834, pRARE2, pVP-16	
Structure by X-ray	Resolution: 2.32 Å	R-value (R-free): 21.0% (23.9%)
	No. of Residues/ASU: 224 (340)	Complexes/ASU: 2
Data Collected At	Advanced Photon Source SER-CAT 22-ID 13-Jun-2005	
Authors	E. Bitto, C.A. Bingman, G.E. Wesenberg, G.N. Phillips, Jr.	



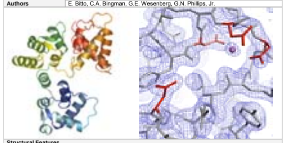
Structural Features

This protein shows 75% sequence identity to human XTP3-transactivated protein A (XTP3TPA, Unicode Q9H773) also known as RS21C6 in other species. Recent structure-guided analysis uncovered a relationship between all-alpha NTP pyrophosphohydrolases (dimeric UTPase, NTP pyrophosphorylase MagZ and RS21C6). Our structure establishes that these proteins share homologous folds and analogous dimeric interfaces. Sequence alignments of these proteins reveals several high conserved negatively charged side chains, which are involved in the coordination of up to three Mg++ ions, and a lysine residue which stablizes the γ-phosphate of NTPs. All these residues (Glu63, Glu66, Glu95, Asp98, and Lys121) are conserved in mouse RS21C6. It has been hypothesized that members of this family may be involved in the housekeeping of non-canonical trinucleotides that could otherwise be incorporated into DNA or RNA. In the case of RS21C6, the protein is hypothesized to act on 5-methyldeoxycytidine. *Reference:* Moroz, O.V. *et al J. Mol. Biol.* (2005) **347**, 243-255.

Percent Identity with Nearest PDB Structure at Time Solved	25% (1VMG)
Pfam Cluster	Pfam-B_61816,5032,40152,72040
Sequence Cluster Size : Structures in PDB	77 at e<0.1

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	74073	
Organism	*Danio rerio*	
Gene Designator	Dr.36843	
PDB Entry	2BE4	Deposition: 21-Oct-2005
Function	EF-hand protein from *Danio rerio*	
Produced From	*E. coli* B834, pRARE2, pVP-16	
Structure by X-ray	Resolution: 2.10 Å	R-value (R-free): 17.6% (25.3%)
	No. of Residues/ASU: 270 (271)	Complexes/ASU: 1
Data Collected At	Advanced Photon Source SBC 22-ID 10-Oct-2005	
Authors	E. Bitto, C.A. Bingman, G.E. Wesenberg, G.N. Phillips, Jr.	



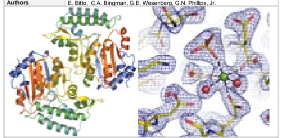
Structural Features

Dr.36843 encodes a novel hexa-EF-hand protein, showing 70% sequence identity to a human protein named "secretogorgin" (Uniprot 076038) Human secretagorgin is preferentially expressed in two tissues: islet cells of the pancreas and neuroendocrine cells. Analysis of the histochemical staining pattern of human secretagorgin revealed a striking neuron-specific cerebral expression pattern. Interestingly, the protein is detectable in human serum after ischemic strokes. This structure of the calcium-free form of D. rerio putative-secretagorgin reveals three independent domains, each of which contains a pair of EF-hand motifs, connected by short linkers.

Percent Identity with Nearest PDB Structure at Time Solved	1TCF (26% over 164 aa)
Pfam Cluster	Efhand (6) Dockerin_1, CMAS
Sequence Cluster Size : Structures in PDB	750 non-redundant genes at e<0.1

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	36414	
Organism	*Mus musculus*	
Gene Designator	NT5C3, Mm.158936	
PDB Entry	2BDU	Deposition: 20-Oct-2005
Function	cytosolic 5´-nucleotidase III	
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.	



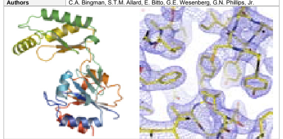
Structural Features

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.

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)
Protein Structure Data Summary

CESG ORF No.	36653	
Organism	*Homo sapiens*	
Gene Designator	Hs.459855	
PDB Entry	2AMY	Deposition: 10-Aug-2005
Function	phosphomannomutase 2 (PMM2)	
Produced From	*E. coli* B834, pRARE-2, pVP-16	
Structure by X-ray	Resolution: 2.09 Å	R-value (R-free): 19.9% (27.2%)
	No. of Residues/ASU: 240 (245)	Complexes/ASU: 1
Data Collected At	Advanced Photon Source SBC 22-ID 13-Jun-2005	
Authors	C.A. Bingman, S.T.M. Allard, E. Bitto, G.E. Wesenberg, G.N. Phillips, Jr.	



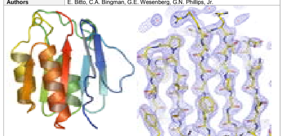
Structural Features

The structure of phosphomannomutase-2 from *Homo sapiens* was determined by SeMet MAD at SER-CAT (APS 22-ID). This enzyme (EC 5.4.2.8) is involved in the synthesis of GDP-mannose and dolichol-phosphomannose required for a number of critical mannosyl transfer reactions. The structure reveals a two domain structure, with the larger, lower domain having a typical hydrolase fold. A DALI search reveals structural similarity to members of the haloacid dehalogenase (HAD) superfamily. PMM2 is known to be involved in a human disease state. Patients with carbohydrate-deficient glycoprotein syndrome type 1 (CDG1 or Jaeken's disease) show a deficiency in PMM activity. This inherited defect usually manifests itself as a severe neonatal disorder, with 20% mortality in the first year of life, due to infections, hepatic insufficiency and cardiomyopathy. On a molecular level, these defects are characterized by defective protein glycosylation.

Percent Identity with Nearest PDB Structure at Time Solved	16% (1NF2)
Pfam Cluster	Hydrolase_3, PMM
Sequence Cluster Size : Structures in PDB	104 NR at e<0.1

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	33759	
Organism	*Homo sapiens*	
Gene Designator	Hs.95870	
PDB Entry	2AB1	Deposition: 14-Jul-2005
Function	unknown	
Produced From	*E. coli* B834, pRARE2, pVP-16	
Structure by X-ray	Resolution: 2.59Å	R-value (R-free): 19.3% (25.7%)
	No. of Residues/ASU: 234 (244)	Complexes/ASU: 2
Data Collected At	Advanced Photon Source SER-CAT 22-ID 13-Mar-2005	
Authors	E. Bitto, C.A. Bingman, G.E. Wesenberg, G.N. Phillips, Jr.	



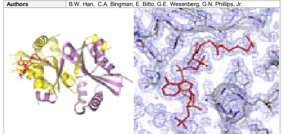
Structural Features

The crystal structure of Hs.95870 revealed a fold seen in two other proteins: MTH938 from *Methanobacterium thermoautotrophicum* and PH1505 from *Pyrococcus horikoshii*. Hs.95870 adopts a three layer sandwich architecture with a central mixed 5-stranded beta sheet, a layer of alpha helices, and a 2-stranded beta sheet. Profile-profile alignments of the three structural homologs reveal several conserved residues. Of particular note is a cluster of His45, His119, Cys122, Thr121, Leu75, and Gly70. The first three of these residues appear to be poised to coordinate a metal.

Percent Identity with Nearest PDB Structure at Time Solved	26% (1IHN)
Pfam Cluster	DUF498 (6-119) CoA_trans (56-88)
Sequence Cluster Size : Structures in PDB	28 NR at e<0.1

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	36731	
Organism	*Homo sapiens*	
Gene Designator	Hs.SSAT-2	
PDB Entry	2BEI	Deposition: 24-Oct-2005
Function	thialysine N-acetyltransferase	
Produced From	*E. coli* B834, pRARE2, pVP-16	
Structure by X-ray	Resolution: 1.84 Å	R-value (R-free): 20.4% (24.9%)
	No. of Residues/ASU: 318 (338)	Complexes/ASU: 2
Data Collected At	Advanced Photon Source SER-CAT 22-ID 10-Oct-2005	
Authors	B.W. Han, C.A. Bingman, E. Bitto, G.E. Wesenberg, G.N. Phillips, Jr.	



Structural Features

Homo sapiens thialysine N-ε-acetyltransferase has been annotated as Hs spermidine N1-acetyltransferase 2 (HsSSAT2) because of high sequence identity to HsSSAT1. However, recent biochemical studies have shown that this protein fails to perform acetyl transfers to polyamines and that thialysine is a substrate for this enzyme. Thialysine is a structural analog of L-lysine, and is known to play a role as an anti-metabolite by competing with lys for incorporation into polypeptides. A high thialysine to lysine ratio has been shown to block the growth of eukaryotic cells. In addition to this anti-metabolite function, thialysine has been identified in mammalian tissues including brain. Thialysine can be converted into cyclic ketimine and other compounds postulated to serve neurochemical roles. This structure is unique, in that it is the only acetyltransferase in the PDB with acetyl-CoA bound to one subunit of a homodimeric acetyltransferase.

Percent Identity with Nearest PDB Structure at Time Solved	17% (1TIQ)
Pfam Cluster	Acetyltransf_1
Sequence Cluster Size : Structures in PDB	636 proteins at e<0.1

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