

Center for Eukaryotic Structural Genomics

Protein Structure Initiative

Structural Genomics of *Arabidopsis* Proteins

Craig A. Bingman

Center for Eukaryotic Structural Genomics
2005-11-18

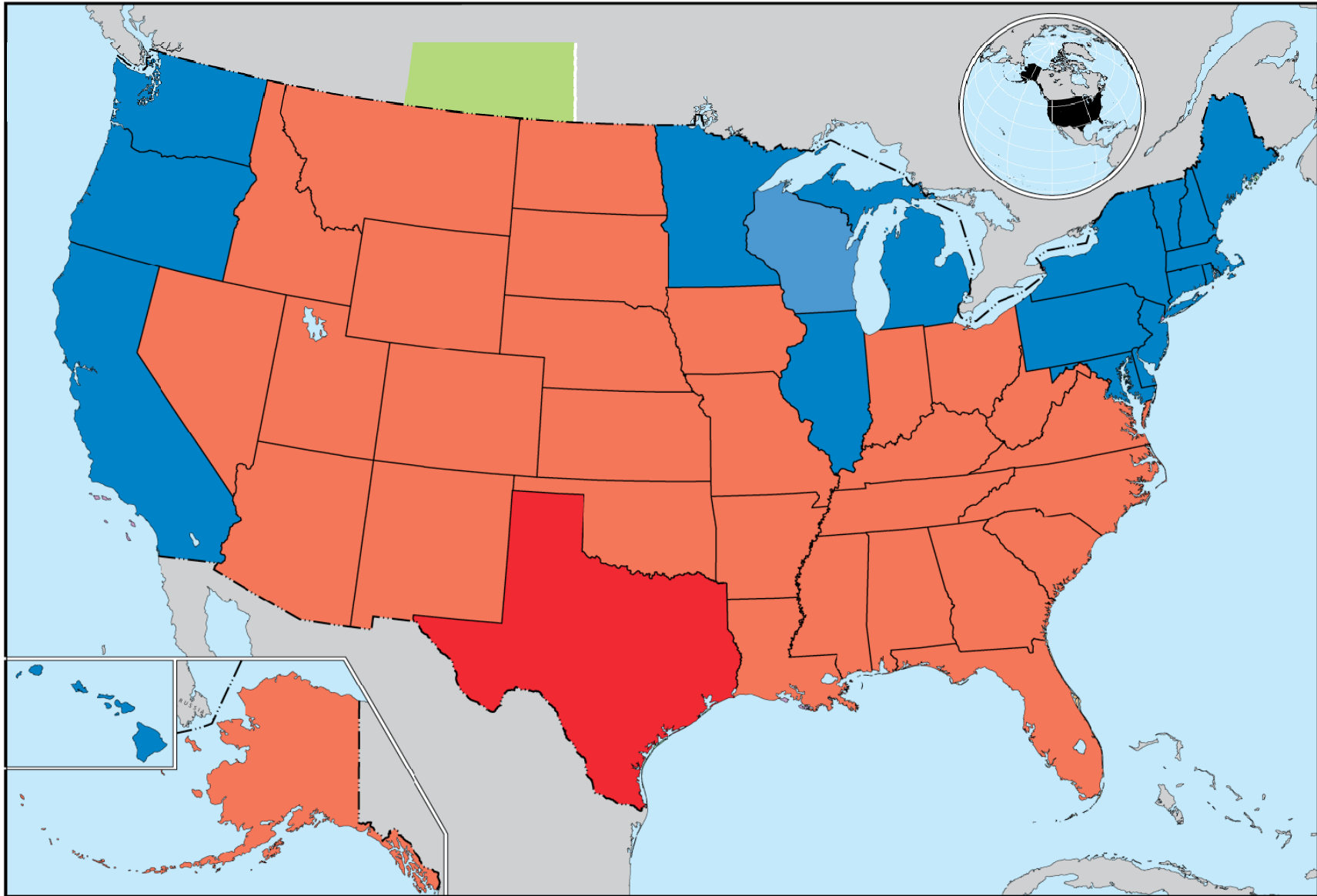


Craig A. Binaman 2005



Center for Eukaryotic Structural Genomics

Protein Structure Initiative



A Brief History of CESH

- 2000 - 7 PSI Pilot Projects Awarded
 - Mandate: explore sequence-structure relationships
- 2001 - 2 Additional Pilot Projects Awarded
- 2002 - CESH creates stable cDNA reagents from *Arabidopsis*
 - No real pipeline, scattered, poorly documented experiments.
- 2003 - Establishment of flow through pipeline
 - Autoinduction medium
 - Gateway cloning
 - Predictive small scale testing
 - Semi-automated purification
- 2004 - Structure Solution Pipeline Ramps Up to 30 structures/yr pace
 - *Arabidopsis* genome effectively mined out, pan-eukaryotic targets
 - Crystal Farm imaging system
 - LIMS drives Tecan optimizations
 - Improved predictive power of small-scale testing
 - Cell-free expression system reaches production status
- 2005 - CESH Awarded Specialized Center for PSI Phase-2
 - Microfluidics crystallization
 - Micro-scale purification
 - Second Crystal Farm

Summary of CESH activities in *Arabidopsis*

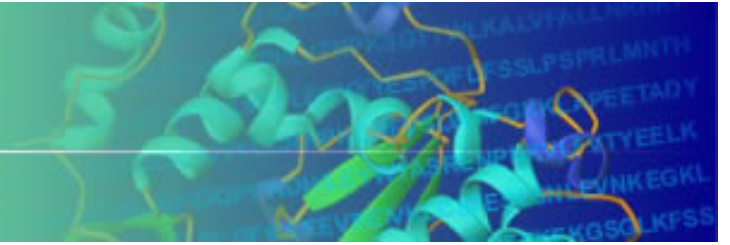
Activity	Instances
Selected Targets	4128
PCR+	3012
Entry Clone+	2526
Sequence+	2434
Destination Clone+	2229
Screening Expression+	1477
Screening Soluble+	1308
Large Scale Cell Growth+	1227
Production Scale Expression+	1132
Soluble Production+	969
Tag Cleaved+	673
Successful Purification	309

Activity	Instances
To Crystallization Screening	237
Crystallized+	97
Diffraction Quality Crystals	42
Crystal Structure+	33
X-ray PDB Deposited	33
HSQC Trials	105
HSQC+	34
NMR Assigned	16
NMR Structure	16
NMR PDB Deposited	16
BMRB Deposited	16

Most recent estimate of Arabidopsis genome size is 26,751 genes encoding proteins.

Over ~4 years, CESH covered 15% of the Arabidopsis genome.

Expression clones were created for 8% of Arabidopsis proteins.



- Arabidopsis clones available

<http://www.uwstructuralgenomics.org/presentations/>

arabidopsis_05_newman_poster.pdf

arabidopsis_05_targetlist.xls

- Active outreach program

- Entry point DNA, sequence data

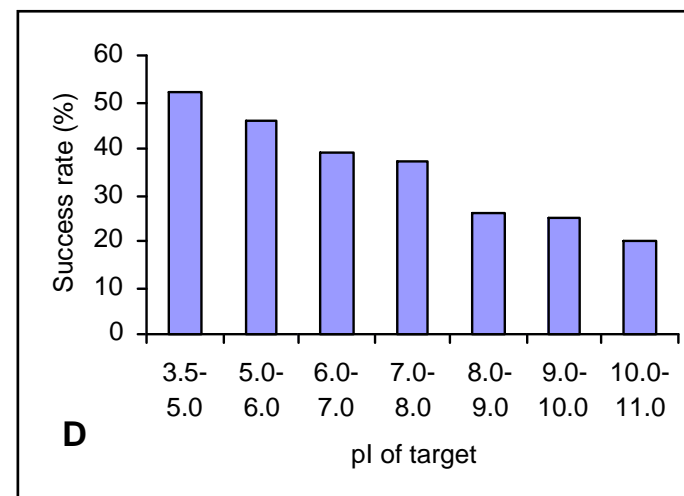
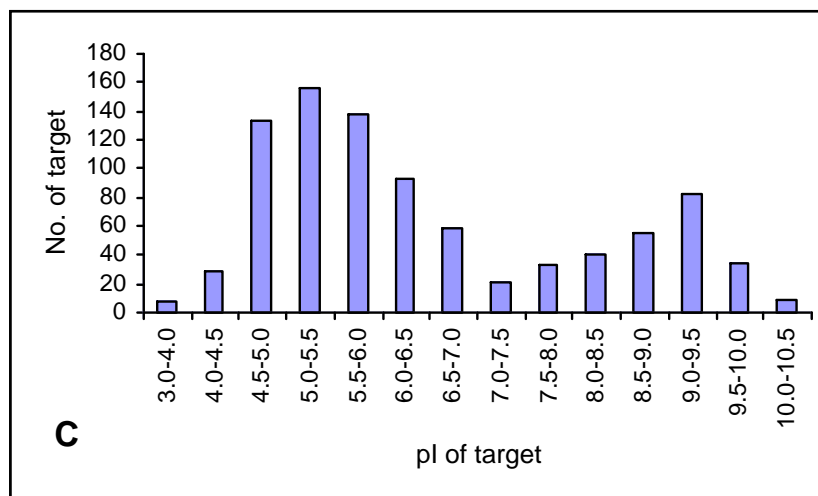
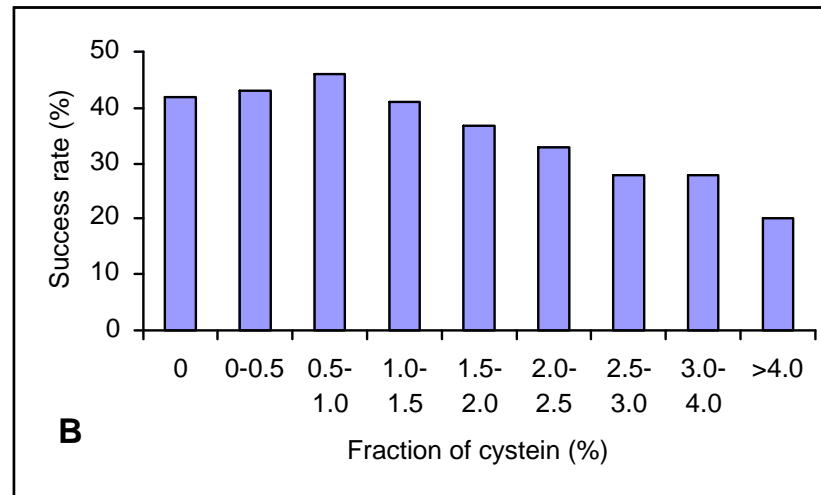
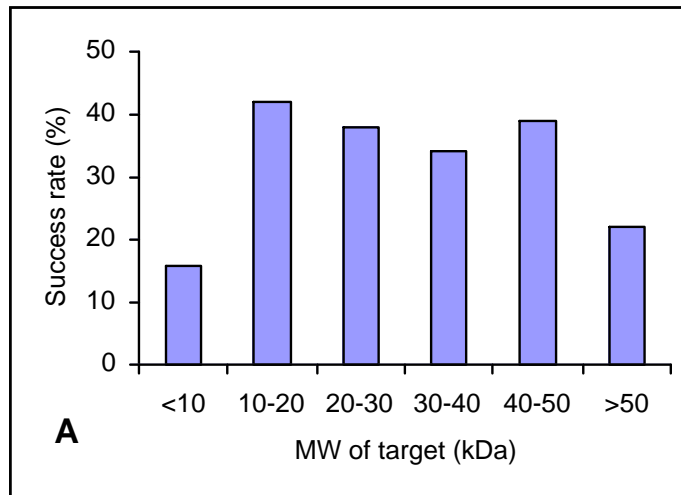
- Entry point purified protein, sequence data

- Contact person, John Primm

<http://www.uwstructuralgenomics.org/structurerequest.htm>



Performance of *Arabidopsis* Proteins, By Calculated Property



How it works

- Pipeline adapted to first-pass SeMet for crystallography samples.
- Ideally, one large-scale sample, screen, optimize, solve structure.
 - Each large scale growth and purification costs \$2,000-3,000.
 - Latency for regrowth request 2 months.
- Deposit structure of SeMet protein, no sample of native protein grown.
- Of the 33 Arabidopsis structures solved to date:
 - 7 MR
 - 1 native metals (Fe/Zn)
 - 1 Br- soak
 - 1 thimerosol (Hg) soak
 - All others, SeMet MAD or SAD
- 24 hours of beamtime = 6 PDB depositions.
- Salvage screening, protein modification helpful.
 - Reductive methylation
 - Organic additives

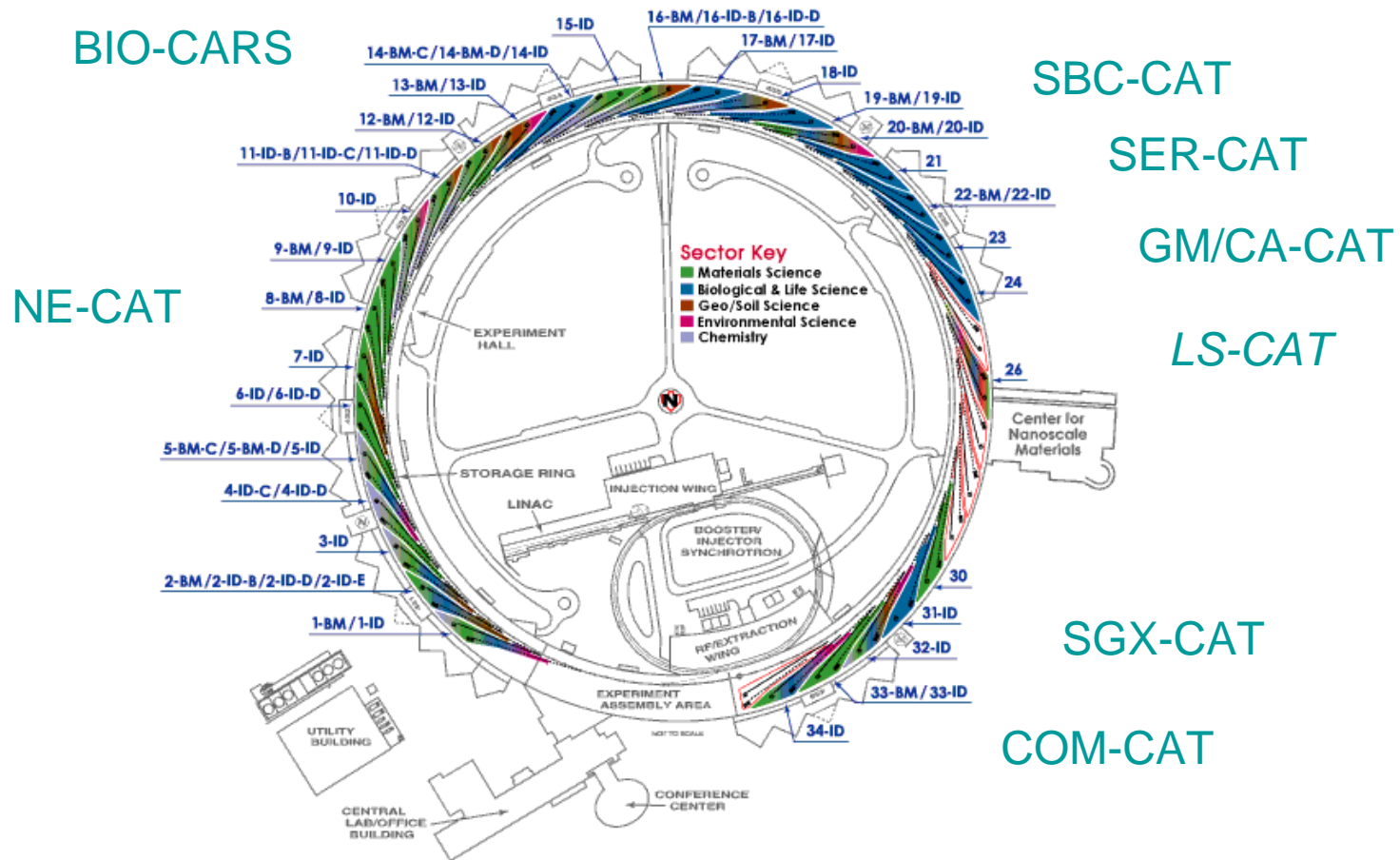
Structure Solution Pipeline

- SeMet MAD or SAD dataset
- Se substructure solved with *phenix.hyss* or *SHELXD*
- Phasing in *SHARP/autosharp*
- Autotracing in *ARP/wARP*
 - *SOLVE/RESOLVE* still useful for cases with NCS.
- Refinement in *refmac5* (TLS refinement increasingly important.)
- Validation using *PROCHECK*, *WHATIF*, *MolProbity*
- Post-structure solution bioinformatics
- Semi-automated PDB deposition
- Semi-automated first draft of publication

Center for Eukaryotic Structural Genomics

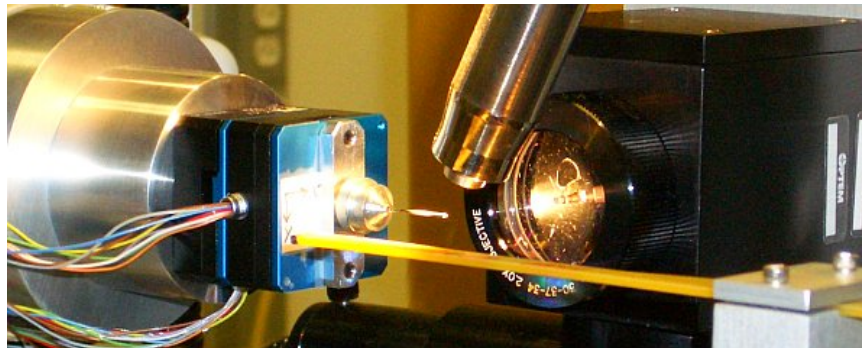
Protein Structure Initiative

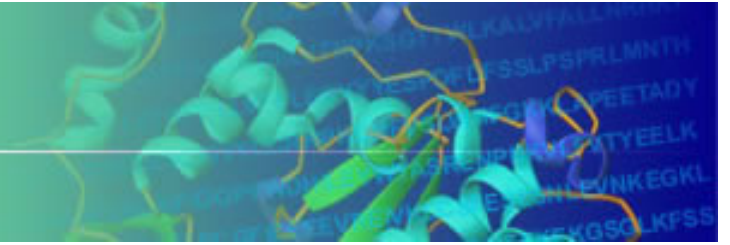
APS Beamlines Used by CESG



Desirable Beamline Attributes

- Fluorescence scans and energy changes in user-space.
- High energy resolution/reproducibility (SER-CAT $\langle 220 \rangle$ Si mono.)
- Efficient fluorescence scans (limit pre-experiment dose to crystal.)
- Fluorescence emission spectrum accessible to user.
- Small “sphere of confusion.”
- Small beam divergence (observed mosaicity of 0.08° at GM/CA-CAT)
- Problems: huge detectors lead to huge downstream data storage liability. (Bin down on demand.)
- Cryostreams. Liquid nitrogen fills still occasionally kill crystals.
- Size of frames beginning to challenge the internet.

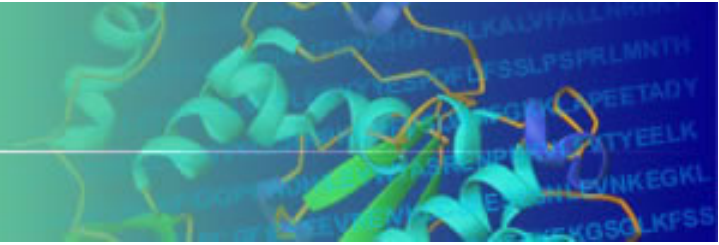




go.29544 At5g21482 CKX

- Target properties

- 524 aa
- 58,304 MW
- pI 4.8
- 8 internal methionines
- Annotated as “putative cytokinin oxidase” in TAIR
- Target selection tier 3, final tier 9
 - 1W1S Released 26 August 2004
- Gel filtration indicates sample is monomeric (“57 kD”)
- Target traversed cell-based expression, single pass
- Obvious yellow color, absorption maxima at 370 and 450 nm
- SeMet incorporation?, QA calls ligand or modification +850Da (FAD +812Da)

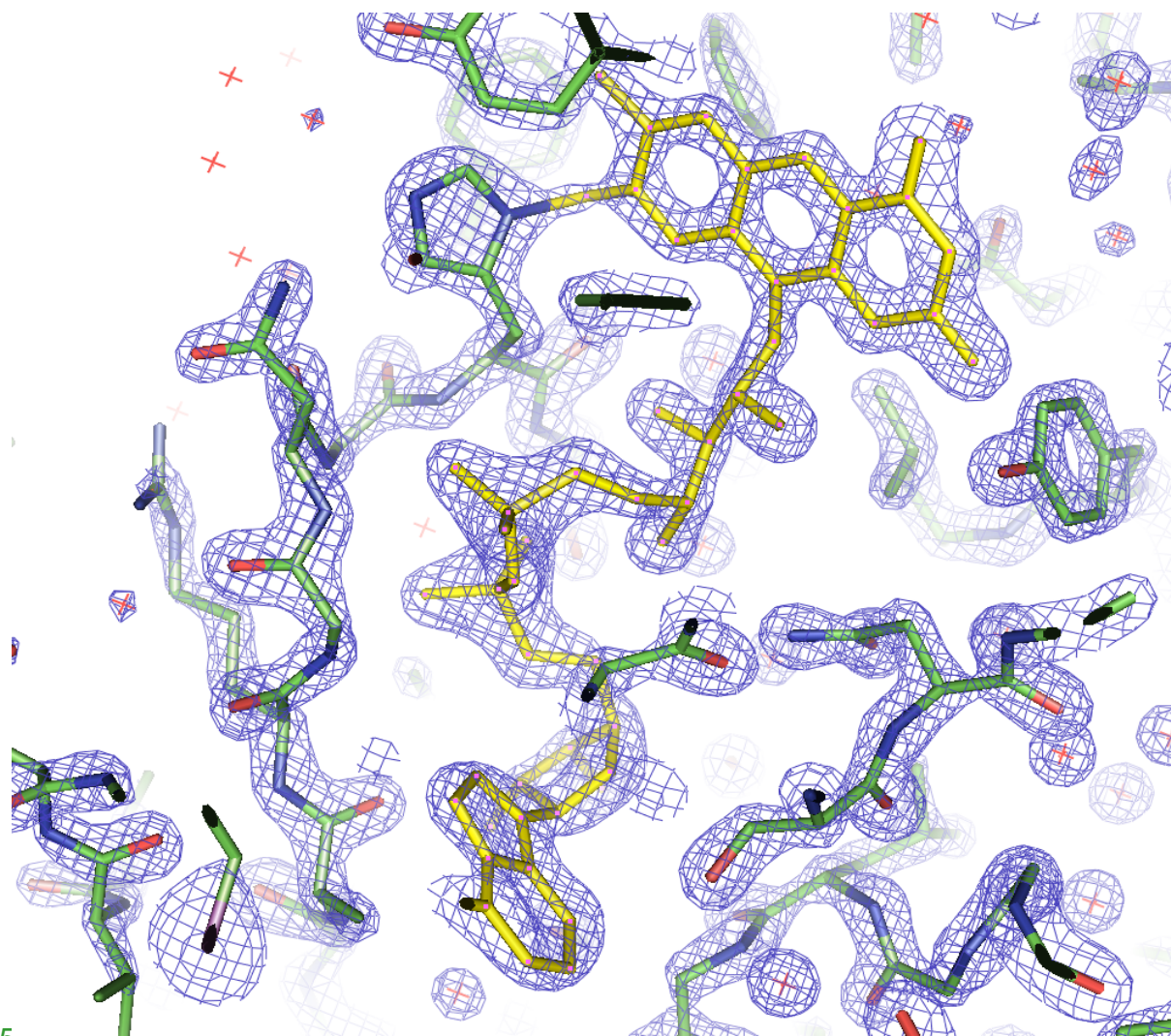


go.29544 At5g21482 CKX 2EXR

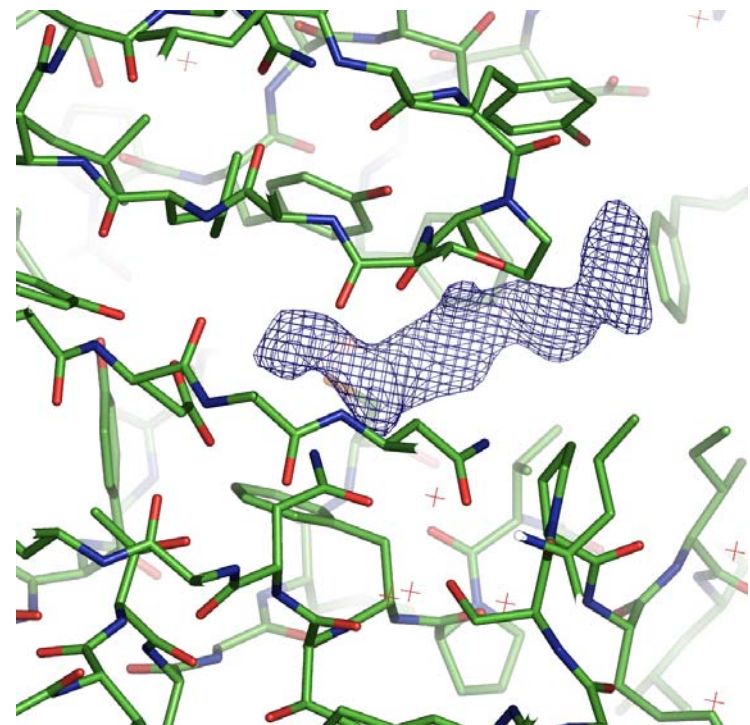
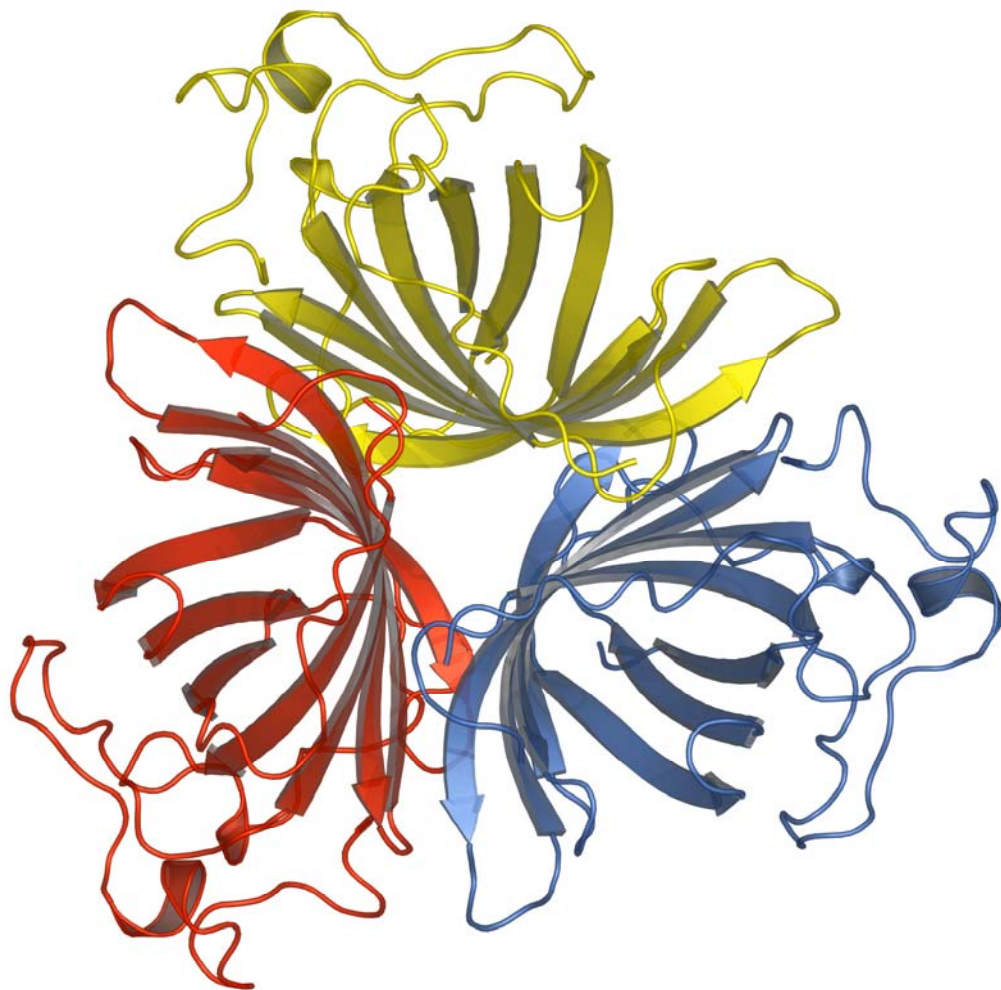
- **Action trace**

• Selected	2004-04-09	csnewman	
• Primers Ordered	2004-04-09	csnewman	
• PCR-	2004-05-18	tkimball	
– 1W1S released	2004-08-26		
• Selected	2004-11-09	csnewman	
• PCR+	2004-12-06	esteffen	
• Entry clone acceptable	2004-12-14	esteffen	
– Homology noted in LIMS	2005-03-03		
• Screening acceptable	2005-03-05	johnk	
• ESC MMH	2005-03-17	btburns	
• Paste acceptable	2005-03-22	fcvojtik	
• Purification complete	2005-04-15	breanne	2670
• QA Data Entered	2005-04-13	gsabat	
• QA Requested Sequence	2005-06-08	daceti	
• QA Activated	2005-07-21	daceti	2550
• Screen Set Up	2005-08-16	zteggers	2110
• Crystallized+	2005-09-01	zteggers	
• Phasing Data +	2005-10-17	bae	1680
• X-ray PDB Deposited	2005-11-08	wesenberg	-990(560)

At5g21482 AtCKX7 FAD Electron Density



At3g25760 Allene Oxide Synthase, Part of Jasmonate Synthesis Pathway



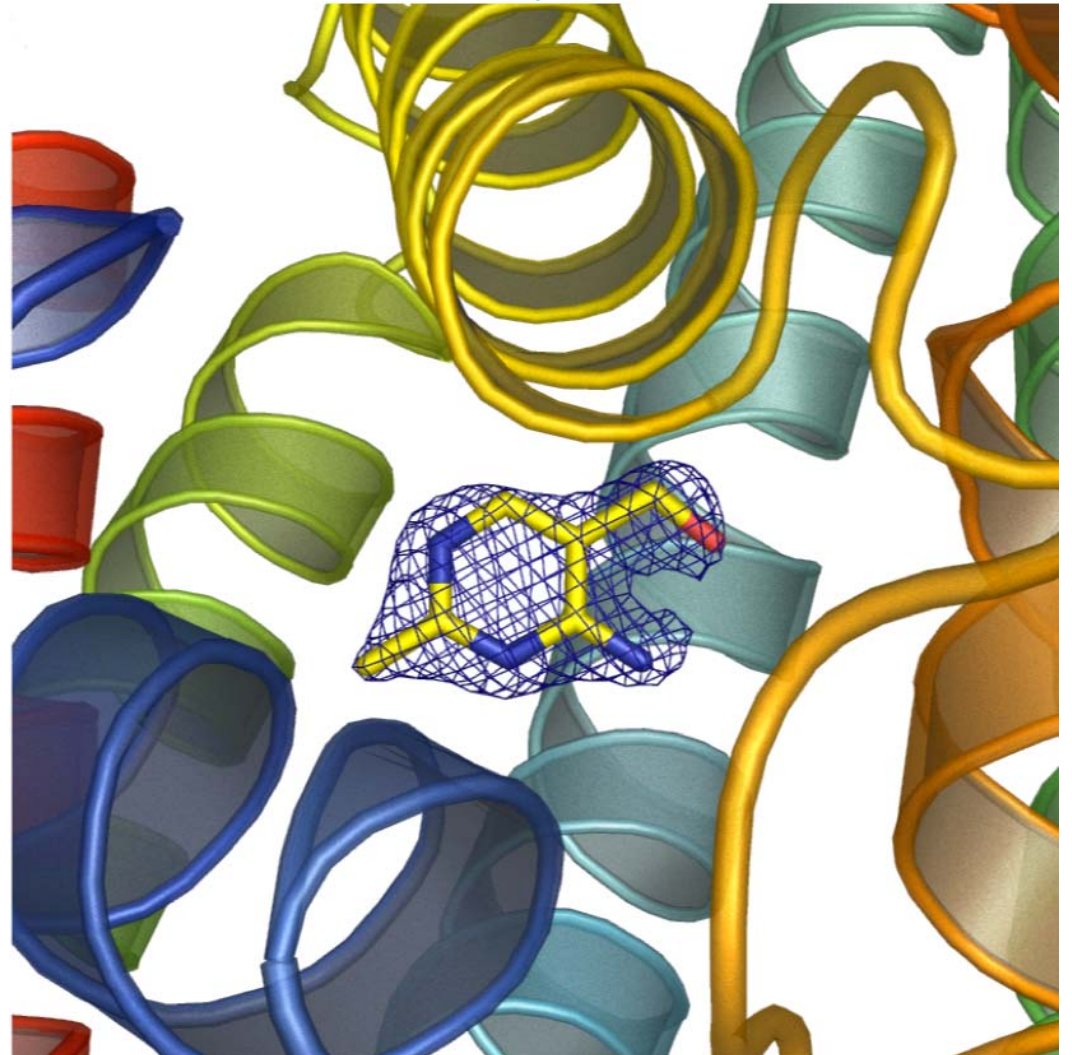
At3g19660: Probable role in thiamine biosynthesis

First CESTG structure.

NESG solved prokaryotic homolog.

Bound 4-amino-5-hydroxymethyl-2-methyl pyrimidine phosphate.

At3g19660 appears to bind 4-amino-5-hydroxymethyl-2-methyl pyrimidine.



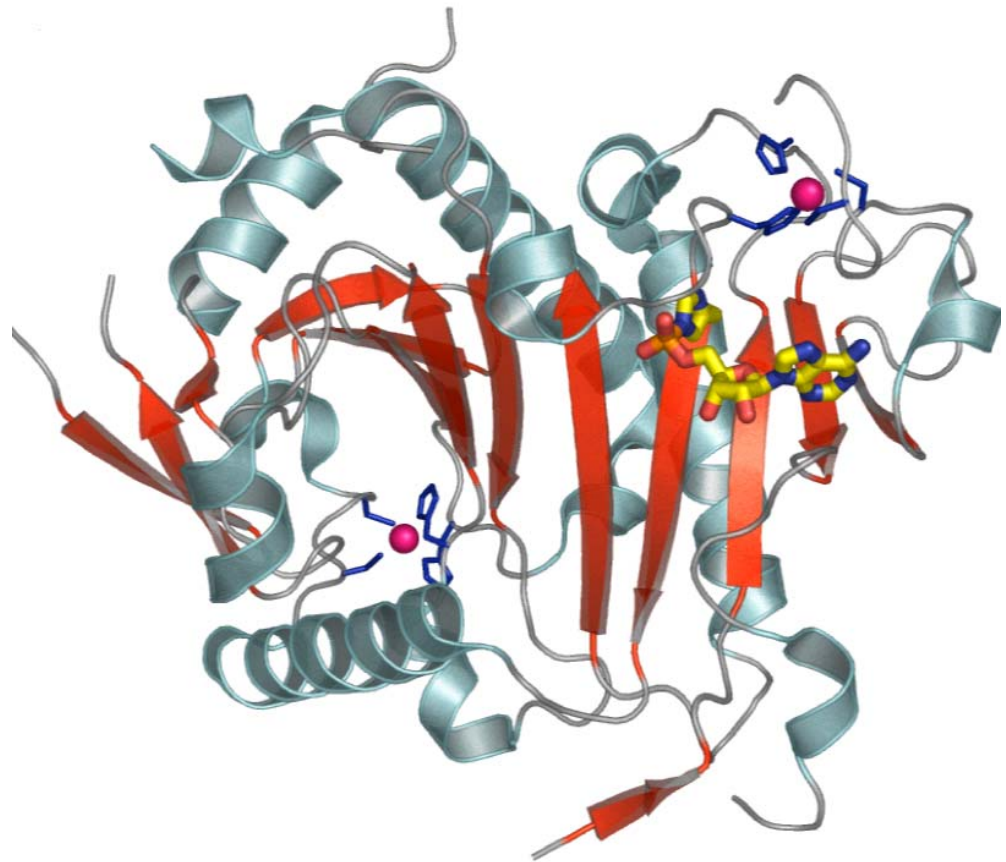
At5g18200: an unknown ADP-glucose pyrophosphorylase

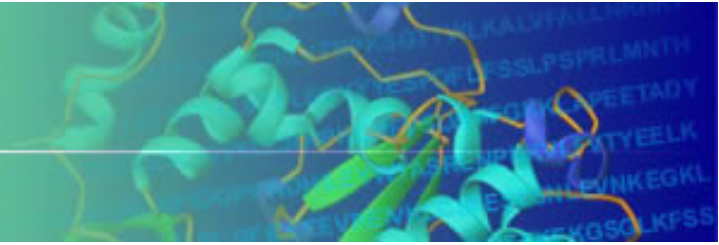
At5g18200 mis-annotated as
UDP-galactose transferase.

Collaboration with Perry Frey
showed it to be 10^4 more active
with ADP-glucose.

Probably involved in ADP-glu
homeostasis.

May play a role in starch
metabolism.





CESG Crystallography/Bioinformatics/Computing Core

George N. Phillips, Jr.

Eduard Bitto, Ph.D.

Euiyoung Bae, Ph.D.

Byung Woo Han

Jason McCoy

Gary Wesenberg, Ph.D.

Zhaohui Sun, Ph.D.

Bryan Ramirez

David Smith, Ph.D.

Simon Allard, Ph.D.

Zach Eggers

Janelle Warrick

Nathan Rosenberg

Jeff Shaw

Amanda Hibbard

Theresa Filarsky

Brandon Wanless

Tony Kamenick

P.I.

crystallography

crystallography

crystallography

crystallography

bioinformatics

bioinformatics

computing core

crystallography

crystallography

crystallomics

crystallomics

crystallomics

crystallomics

undergrad, crystallomics

undergrad, crystallomics

undergrad, crystallomics

undergrad, computing core

CESG *Dramatis Personae*

<http://www.uwstructuralgenomics.org>

John Markley	P.I.	Betsy Lytle	NMR
George N. Phillips	co-P.I.	Emily Peterson	Admin
Brian Fox	co-P.I.	Francis Peterson	NMR
Brian Volkman	EC	Megan Ritters	Cloning
Simon Twigger	EC	Grzegorz Sabat	Mass Spectrometry
John Primm	PM	Kory Seder	Cloning
		Jikui Song	NMR
		Hassan Sreenath	<i>E. coli</i> Expression
David Aceti	Quality Assurance	Donna Troestler	Admin
Paul Blommel	Vector Development	Robert Tyler	NMR
Brendan Burns	<i>E. coli</i> Expression	Dmitriy Vinarov	Cell-free
Claudia Conilescu	NMR	Frank Vojtik	Protein Production
Jason Ellefson	Protein Purification	Jeanette Waltner	NMR
Ronnie Frederick	Expression Testing	Rita Warden	Admin
Won Bae Jeon	Protein Purification	Liya Wang	NMR
Peter Lee	Sesame	Russell Wrobel	Cloning
Carrie Newman	Cell-free	Zsolt Zolnai	Sesame
<i>Craig Newman</i>	<i>Sesame/Bioinfo.</i>		