

Supplemental Update: Structural Genomics of Eukaryotic Domain Families GM-U54-094597

Structure Production Goals. In Yr 5 (through Feb 28, 2010), NESG has deposited 91 structures to the PDB, of which 81 are 'Distinct Structures'. These 36 NMR structures and 55 X-ray crystal structures include 9 human protein structures. Many more structures are in final stages of data collection or refinement. We expect to deposit 180 – 200 structures by Jun 30, 2010; the final structure count from PSI1+PSI2 to be ~ 950 by Sept 1, 2010.

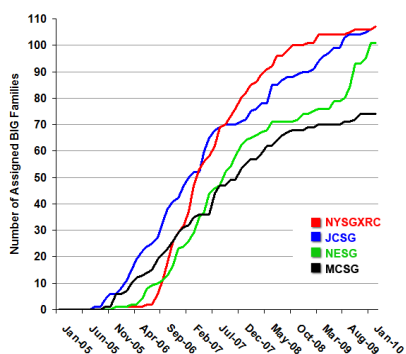


Fig. 1. Progress on BIG Families

by providing 42 new Structures for 37 Distinct Targets to the PDB. Fig. 2 illustrates the success by NESG on BIG targets compared with the other three LSCs. Fig. 2A summarizes the *number of BIG families* covered by each center that were originally assigned to the LSC (yellow), together with the number of BIG families covered by each center that were originally assigned to another LSC (blue) (as explained in the original proposal). NESG has provided structures for 129 BIG families; 101 assigned to NESG (yellow), as well as for an additional 28 assigned to other LSCs. Fig. 2B summarizes the *number of structures* from BIG families assigned to each LSC (yellow) or to another LSC (blue); in each case the left bar counts Total Structures, and the right bar Distinct Structures; e.g. NMR / Xray pairs of the same protein count as two structures in the left bar, and as 1 Distinct Structure in the right bar. NESG has deposited some 236 Total Structures (206 Distinct) from BIG families, more than any other LSC, and some 183 Total Structures (159 Distinct Structures) from BIG families assigned to NESG.

BIG (pfam) Families. Coverage of "BIG" (which include pfam) domain families was an important goal of PSI2, and efforts to expand coverage of the BIG domain families will continue in PSI:BiologY. For these reasons, NESG has made a renewed effort in Yr 5 to salvage targets from BIG families so as to expand our coverage of these families. As illustrated in Fig. 1, plotting

coverage of BIG families assigned to each Large Scale Center (LSC), this ongoing salvage effort is very successful. Initiating efforts to salvage "Open BIGs" in Mar 09, we have increased coverage of 604 BIG

families assigned to NESG from 75 to 101 families, by providing 42 new Structures for 37 Distinct Targets to the PDB. Fig. 2 illustrates the success by NESG on BIG targets compared with the other three LSCs. Fig. 2A summarizes the *number of BIG families* covered by each center that were originally assigned to the LSC (yellow), together with the number of BIG families covered by each center that were originally assigned to another LSC (blue) (as explained in the original proposal). NESG has provided structures for 129 BIG families; 101 assigned to NESG (yellow), as well as for an additional 28 assigned to other LSCs. Fig. 2B summarizes the *number of structures* from BIG families assigned to each LSC (yellow) or to another LSC (blue); in each case the left bar counts Total Structures, and the right bar Distinct Structures; e.g. NMR / Xray pairs of the same protein count as two structures in the left bar, and as 1 Distinct Structure in the right bar. NESG has deposited some 236 Total Structures (206 Distinct) from BIG families, more than any other LSC, and some 183 Total Structures (159 Distinct Structures) from BIG families assigned to NESG.

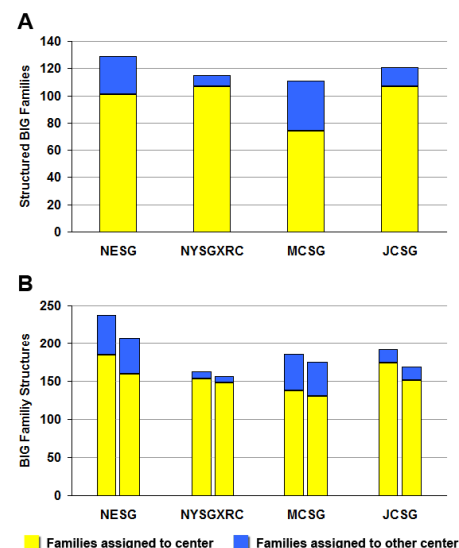


Fig. 2. Progress on BIG Families

Human Cancer Protein Interaction Network (HCPIN). Progress in our HCPIN cancer biology project is

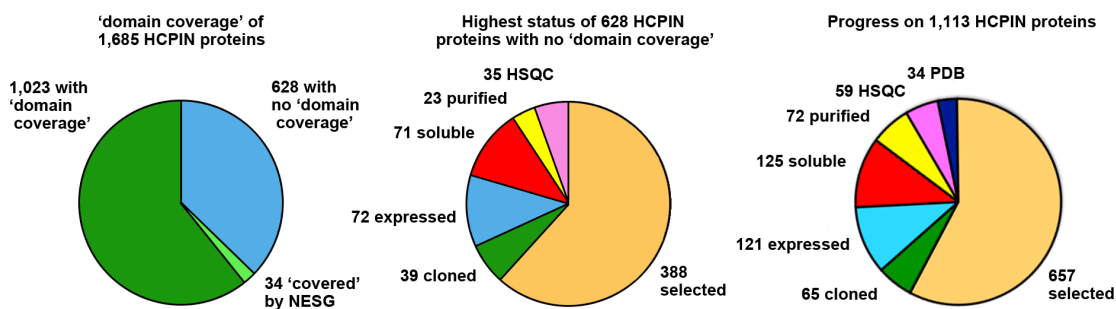


Fig.3. Left. "Domain Coverage" (defined in text) of 1,685 intracellular human cancer proteins of the HCPIN Network. **Middle.** Most advanced status on 628 HCPIN proteins with no "Domain Coverage" **Right.** Most advanced status for 1133 HCPIN proteins 'in progress' in the NESG project.

summarized in the charts of Fig. 3. These large human proteins generally have multiple domains. "Domain coverage" is defined as "structure determination of one or more domains of a multidomain protein (or of a homologous domain with > 80% seq_id)". Thirty-six (36) NESG human protein structures have provided new "domain coverage" (first structure for part of an HCPIN protein) for 34 human cancer-associated proteins (including three secreted proteins); six (6) HCPIN structures have been deposited since Oct 15; e.g. CDK2-AP1 (p12^{DOC-1}), a well known tumor suppressor protein. The most advanced status of 628 human cancer protein targets with no "domain coverage" is summarized in the middle panel; constructs have been designed for all 628; ~ 40% of these have been cloned to date. We expect to complete cloning/expression analysis of at least one domain from these 628 HCPIN proteins by Jun 30. The right panel summarizes progress on 1133 HCPIN proteins for which constructs have been designed, including the 628 with no "domain coverage" and 505 for which one or more domains already have structures. This project will continue in the PSI:BiologY program.

Human Ub Mega Domain Family. Progress on the human Ubiquitin-Like (UBL) Mega family, our primary Mega family objective in Yr 5, is summarized in Fig. 4. As outlined in our proposal, 11 UBL Modeling Families (MFs)

have been covered by NESG and 29 remain "Open" (Fig. 4, left). Sixty-four (64) protein domains, including one or more members of each of the 29 "Open" human UBL MFs, have now been cloned. Seven (7) of these domain targets have provided soluble samples, some of which are in NMR and/or crystallization screening.

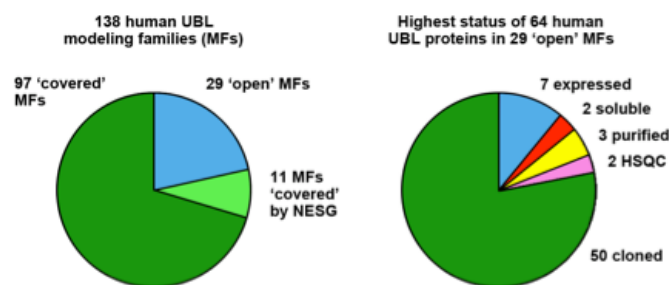


Fig. 4. Progress on human UBL Domain MFs

each of 126 Open MFs have been cloned, and the most advance status each of these MFs is illustrated in Fig. 5. As an example, the solution structure of the *B. subtilis* protein YndB, a salvage project protein that was not successfully crystallized by the MCSG, has been solved using NMR (Mercier, et al. 2010). The protein exhibits the typical helix-grip fold of the Bet v 1-like superfamily (which includes START domains). This fold forms a hydrophobic cavity that binds lipid-like molecules. An *in silico* screen of a compound library of ~18,500 lipids was used to identify classes that preferentially bind YndB. The screen identified chalcone /

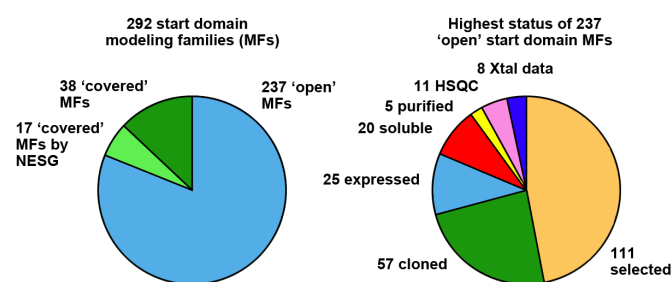


Fig. 5 Progress on START Domain MFs

START Mega Domain Family. We have applied the same bioinformatics methods outlined in the proposal for defining UBL MFs in the human genome to define 292 MFs for START domain Mega family. This more rigorous analysis, spanning ~ 180 NESG Reagent Genomes, reveals that comprehensive coverage of the START domain Mega family will require at least 237 new 3D structures, much more than originally anticipated. To date, NESG has solved 15 structures covering 17 START MFs; providing ~1/3 of the structural coverage of this Mega family. 1100 targets from the remaining 237 Open Start MFs have been selected for the NESG pipeline; at least one target from hydroxychalcone, flavanone, and flavone/flavonol classes of lipids as likely binders. These results of the *in silico* screen were then verified by 2D ^1H - ^{15}N HSQC NMR titration experiments, where trans-chalcone, flavanone, flavone, and flavonol were all shown to bind in the YndB hydrophobic cavity with K_D values ranging from 20 μM to 230 μM . These compounds are often used by plants for many functions, including microbial defense and signaling. The symbiotic relationship between *B. subtilis* and plant roots, along with other genomic data, suggests an involvement of YndB in bacterial stress response to flavonoids released by plants due to pathogen infection.

Interleukin-20 Production in Human HEK293T Cells. Several human proteins that are not well expressed in bacteria or *P. pastoris* systems were tested for expression in human HEK293T cells, using methods outlined in the NESG proposal. Among several clones that expressed well in transient transfections, interleukin-20 (IL-20) was chosen for production in a stable cell line. Transiently transfected 293T cells expressing IL-20 fused to an F_c antibody fragment were put under hygromycin selection for about 2 weeks. Ninety-six stable clones were isolated and tested for IL-20 expression. The top four expressing clones were selected for expansion and a second round of testing. Among those four, the clone with the highest expression signal by anti- F_c Western was seeded into the bioreactor. Several liters of supernatant were produced using the bellows system, and affinity purified using protein A resin. The fusion protein was cleaved with thrombin, liberating IL-20. SDS-PAGE demonstrates a band of approximately 20 kDa (Fig. 6), with a yield of 1.7 mg. MALDI-TOF demonstrates formation of three disulfide bonds. The stable cell line continues to produce IL-20 and is being acclimated for SeMet enrichment.

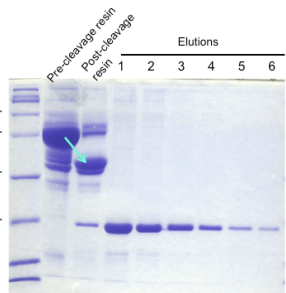


Fig. 6 Production of IL-20 in human HEK293T cells

These results further demonstrate our ability to produce proteins, like IL-20, that cannot be produced in *E. coli* systems using HEK293T cells.

Polycistronic Expression of Multi-protein Complexes. Many proteins function in multi-protein complexes. In many cases the individual proteins that make such complexes may not be tractable for structure determination in the absence of their partners. In our proposal, we described our project to define groups of functionally similar orthologous proteins, called Classes of Reciprocal Sequence Homologs (CRSHs). Data supporting such CRSHs include conservation of gene neighborhoods; i.e. genes coding for proteins that form multi-protein complexes are found in close proximity in divergent prokaryotic genomes. These genes are often co-regulated at the transcriptional level, and the proteins are translated from polycistronic messages. We are taking advantage of this fact to produce large multiprotein complexes, cloning entire or partial operons into our modified T7-based expression systems. This allows several proteins contributing to a multiprotein complex to be translated in a co-cistronic fashion. Our case study involves the Citrate Lyase complex of the 11 kDa CitD, 33 kDa CitE and 55 kDa CitF proteins, that function to convert citrate to oxaloacetate. The ~600 kDa functional complex contains six copies of each of three proteins. Using a pan-genomic approach, we targeted the Cit

complexes from *Streptococcus mutans*, *Salmonella typhimurium*, *Shigella flexneri* and *E. coli*. Operon segments containing all three genes were cloned into either an N-terminal or C-terminal 6X His fusion vector. The former vector adds the tag to CitD, and the latter to CitF, allowing for co-purification all three proteins if they form a complex. Analytical scale expression of the polycistronic messages was performed and the results visualized by SDS-PAGE analysis. All three proteins (CitD, CitE, CitF) from *S. mutans* are expressed as soluble

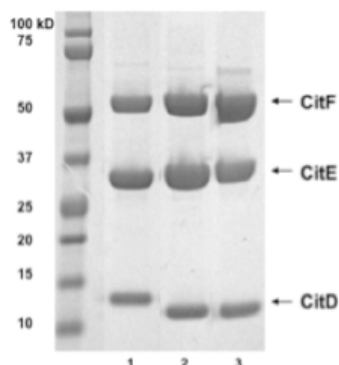


Fig. 7. SDS-PAGE of Citrate lyase complexes

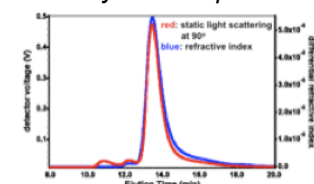


Fig. 8. Monodisperse gel filtration profile for a 600 kDa Cit complex

proteins with either CitD or CitF tagged. The three *S. typhimurium* proteins were also co-expressed in soluble form with a N-terminal 6X His tag on CitD. Stable complex formation was then assessed by Ni-NTA co-purification (Fig. 7). The 6X-His tagged *S. mutans* CitD co-purified with CitE and CitF (lane 1). Likewise, the tagless CitD and CitE co-purified with 6X-His tagged CitF (lane 2), also indicating a stable complex. Similar results were obtained with the 6X-His tagged CitD from *S. typhimurium* (lane 3). Currently, we are purifying these complexes from preparative scale fermentations and screening for crystallization. Fig. 8 shows a monodisperse analytical gel filtration profile for one of these purified 600 kDa complexes. We also cloned *S. mutans* CitD, CitE and CitF into the pDuet system (Novagen), but did not detect expression of all three proteins or complex formation with this commonly used co-expression system. Its failure suggests that some aspect of the polycistronic nature of the Cit operon is important for expression and complex formation. Based on these very promising results, we are identifying other protein complexes from the bacterial CRSH data for polycistronic co-expression and structure determination. Synthetic polycistronic bacterial expression vectors have also been used successfully by NESG investigators to produce *eukaryotic multiprotein complexes*; e.g. the structure of the SNF1 heterotrimer complex by Liang Tong's group (Amodeo, et al. 2007 Nature 449: 492).

NMR Methods Development. NESG NMR methods development has focused on solving structures of homodimers and larger (> 25 kDa) proteins. NESG has determined 3D structures of 25 homodimers by NMR (22 in PSI2), and 1 tetramer. Isotope-filtering strategies for distinguishing intra from intermolecular NOE interactions required for solving dimer structures are used extensively by the NESG NMR groups. However, they are laborious and are not effective in cases of weak dimers, where intermolecular NOEs are exchange broadened and/or very

weak. The Prestegard group has developed a novel procedure for NMR structure determination of *weak homodimers*, combining residual dipolar couplings (RDCs), chemical shift changes upon dilution, and paramagnetic surface perturbations (Lee, et al. 2010). This procedure was applied to NESG protein target, Ser13 from *S. epidermidis*, which forms a weak homodimer (K_D 3.4 mM) at pH 6.5. The structure of dimer was determined by docking subunits, constrained in orientation relative to a C_{2v} axis determined using RDCs data obtained at multiple sample concentrations, and scoring docked geometries using residue pair potentials and shape-based predictions of RDC. The derived model was then validated with paramagnetic surface perturbation data and chemical shift perturbations on sample dilution. The Kennedy group has also made progress, presented at the recent SG Keystone Symposium, using EPR methods, including Double Electron-Electron Resonance (DEER), and Paramagnetic Relaxation Enhancement (PRE) data, to solve the structure of homodimeric NESG target DhR8C, and to measure the kinetics of homodimer chain exchange. Working together with the Baker laboratory, several NESG investigators have contributed to the development of CS-RDC-Rosetta, using RDC data to guide Rosetta calculations. We have demonstrated that good quality structures of proteins up to 25 kDa can be generated with NMR-directed Rosetta calculations using only backbone chemical shift, RDC data, and some NH-NH NOE data, which can be obtained for perdeuterated proteins. This approach, published in Science (Raman, et al. 2010), was developed using NMR data on 2H , ^{13}C , ^{15}N -enriched NESG proteins. SAXS data has been collected by the DeTitta group on 20 proteins for which NESG has solved both NMR and X-ray crystal structures; these data are being used to develop a new approach in which backbone chemical shift and SAXS data are combined to guide Rosetta structure predictions. These hybrid approaches will allow NMR studies of larger (25 – 50 kDa) proteins.

Community Outreach, Materials Repository, Institutional Commitments, and Publications. Since submission of the proposal, we have undertaken three (3) additional community outreach projects, with Profs. C. Clarke (UCLA), B. Kuhlman (UNC-Chapel Hill) and D. West (Fox Chase Cancer Center). We have also completed four (4) community outreach structures (3 by X-ray, 1 by NMR). These include the human Bfl-1:Noxa BH3 peptide complex (C. Gelinas, Cancer Institute of NJ), the GT domain of human mRNA capping enzyme (A. Shatkin, CABM, Rutgers), a redesigned E3 ubiquitin ligase (B. Kuhlman), and a novel lipoprotein (M. Inouye, RW Johnson Med School). NESG also organized a workshop at the recent SG Keystone Symposium highlighting our NMR Technology Wiki site. Extensive bioinformatics efforts have been made to provide fully compliant data for 17,417 unique NESG protein expression vectors that are being provided to PSI-MR, and 128 96-well plates, corresponding to about 5,000 NESG constructs, have been delivered to the PSI-MR (each plate has 2 – 3 isolates of each unique construct; i.e. about 40 unique constructs per plate). We will deliver ~20,000 unique expression vectors to the PSI-MR by Jun 30. Construction on the new Rutgers Proteomics Building is now in progress. Since Oct 15, 2009, 35 NESG papers have been published or submitted for publication.

Supplementary Update: Structural Genomics of Eukaryotic Domain Families GM-U54-094597**Manuscripts Published, Accepted, or Submitted since Oct 15, 2009****PUBLISHED (Formerly “Submitted” or “In Press”)**

1. Petrey D, Fischer M, Honig B (2009) Structural relationships among proteins with different global topologies and their implications for function annotation strategies. **Proc. Natl. Acad. Sci. USA**, 106, 17377-82. Pub Med ID: 19805138; PMCID: PMC2765090.
2. Rosato A, Bagaria A, Baker D, Bardiaux B, Cavalli A, Doreleijers JF, Giachetti A, Guerry P, Güntert P, Herrmann T, Huang YJ, Jonker HR, Mao B, Malliavin TE, Montelione GT, Nilges M, Raman S, van der Schot G, Vranken WF, Vuister GW, Bonvin AM (2009) CASD-NMR: critical assessment of automated structure determination by NMR. **Nature Methods**, 6, 625-6. Pub Med ID: 19718014; NIHMS177629.
3. Schneider WM, Inouye M, Montelione GT, Roth MJ (2009) Independently inducible system of gene expression for condensed single protein production (cSPP) suitable for high efficiency isotope enrichment. **J. Struct. Funct. Genomics**, 10, 219-25. Pub Med ID: 19642019.
4. Mao L, Tang Y, Vaiphei ST, Shimazu T, Kim SG, Mani R, Fakhoury E, White E, Montelione GT, Inouye M (2009) Production of membrane proteins for NMR studies using the condensed single protein (cSPP) production system. **J. Struct. Funct. Genomics**, 10, 281-289. Pub Med ID: 19856129; NIHMS183671.
5. Bromberg Y, Rost B (2009) Correlating protein function and stability through the analysis of single amino acid substitutions. **BMC Bioinformatics**, 10 Suppl 8, S8. Pub Med ID: 19758472; PMCID: PMC2745590
6. Mercier KA, Mueller GA, Acton TB, Xiao R, Montelione GT, Powers R (2009) (1)H, (13)C, and (15)N NMR assignments for the Bacillus subtilis yndB START domain. **Biomolecular NMR Assignments**, 3, 191-4. Pub Med ID: 19888688.
7. Chimnaronk S, Forouhar F, Sakai J, Yao M, Tron CM, Atta M, Fontecave M, Hunt JF, Tanaka I (2009) Snapshots of dynamics in synthesizing N6-Isopentenyladenosine at the tRNA anticodon. **Biochemistry**, 48, 5057-5065
8. Liu G, Xiao R, Wang D, Huang YJ, Acton TB, Montelione GT (2010) Solution NMR structure F-actin binding domain of Arg/Ab12 from Homo sapiens. **PROTEINS: Structure Function Bioinformatics**, 78, 1326-1330. PMCID: PMC2821974.
9. Schaefer C, Schlessinger A, Rost B (2010) Protein secondary structure appears to be robust under in silico evolution while protein disorder appears not to be. **Bioinformatics**, 26, 625-631 Pub Med ID: 20081223; PMCID: PMC2828120
10. Singarapu KK, Mills JL, Xiao R, Acton T, Punta M, Fischer M, Honig B, Rost B, Montelione GT, Szyperski T (2010) Solution NMR structures of proteins VPA0419 from Vibrio parahaemolyticus and yjiS from Shigella flexneri provide structural coverage for protein domain family PFAM 04175. **PROTEINS: Structure Function Bioinformatics**, 78, 779-84. Pub Med ID: 19927321.
11. Raman S, Huang YJ, Mao B, Rossi P, Aramini JM, Liu G, Montelione GT, Baker D (2010) Accurate automated protein NMR structure determination using unassigned NOESY data. **J. Am. Chem. Soc.**, 132, 202-7 Pub Med ID: 20000319.
12. Rossi P, Swapna GV, Huang YJ, Aramini JM, Anklin C, Conover K, Hamilton K, Xiao R, Acton TB, Ertekin A, Everett JK, Montelione GT (2010) A microscale protein NMR sample screening pipeline. **J. Biomolecular NMR**, 46, 11-22. Pub Med ID: 19915800; PMCID: PMC2797623

13. Das R, André I, Shen Y, Wu Y, Lemak A, Bansal S, Arrowsmith CH, Szyperski T, Baker D (2009) Simultaneous prediction of protein folding and docking at high resolution. **Proc. Natl. Acad. Sci. USA.**, 106, 18978-83 Pub Med ID: 19864631; PMCID: PMC2770007.
14. Raman S, Lange OF, Rossi P, Tyka M, Wang X, Aramini J, Liu G, Ramelot T, Eletsy A, Szyperski T, Kennedy M, Prestegard J, Montelione GT, Baker D (2010) NMR structure determination for larger proteins using backbone-only data. **Science**, 327, 1014-8. Pub Med ID: 20133520.
15. Mao L, Vaiphei ST, Shimazu T, Schneider WM, Tang Y, Mani R, Roth MJ, Montelione GT, Inouye M (2010) The E. coli single protein production system for production and structural analysis of membrane proteins. **J. Struct. Funct. Genomics**, epub ahead of print. Pub Med ID: 19830594.
16. Kim D, J.F. Hunt, JF, T. Schirmer, T. (2010) Making, breaking and sensing of cyclic di-GMP: structural aspects. In *The Second Messenger Cyclic Diguanylate*. American Society for Microbiology Press, Washington, DC. (Peer-reviewed review article.)
17. Arragain S, Garcia-Serres R, Blondin G, Douki T, Clemancey M, Latour J-M, Forouhar, F, Neely, H, Montelione, GT, Hunt, JF, Mulliez, E., Fontecave, M, Atta, M. (2010) Post-translational modification of ribosomal proteins: Structural and functional characterization of RimO from *Thermotoga maritima*, a Radical-SAM methylthiotransferase. **J. Biol. Chem.** epub ahead of print. PubMed ID: 20007320 PMCID: PMC28200805

IN PRESS (formally "Submitted")

18. Montelione GT, Szyperski T. (2010) Advances in protein NMR impacting drug discovery provided by the NIGMS Protein Structure Initiative. **Curr Opin Drug Discovery**, In Press.
19. Liu, G, Huang, YJ, Xiao, R, Wang, D, Acton, TB, Montelione, GT (2010) Solution NMR structure of the ARID domain of human AT-rich interactive domain-containing protein 3A: A human cancer protein interaction network target. **PROTEINS: Structure Function Bioinformatics**. In Press.
20. Schneider, WM, Tang, Y, Vaiphei, ST, Mao, L; Maglaqui, M, Inouye, M, Roth, MJ, Montelione, GT (2010) Efficient condensed-phase production of perdeuterated soluble and membrane proteins. **J. Struct. Funct. Genomics**. In Press.
21. Wu B., Skarina T, Yee A., Jobin M.-C., Semesi A., Fares C., Lemak A., Coombes B., Arrowsmith C.H., Singer A.U., Savchenko A. (2010) Nleg type 3 effectors from enterohaemorrhagic Escherichia coli are u-box E3 ubiquitin ligases. **PLoS Pathogens**. In Press.
22. Lee H, Li Z, Silkov A, Fischer M, Petrey D, Honig B, Murray D (2010) High-throughput computational structure-based characterization of protein families: START domains and implications for structural genomics. **J. Struct. Funct. Genomics** In Press

SUBMITTED

23. Aramini JM, Tubbs JL, Kanugula S, Rossi P, Belote RL, Ciccocanti CT, Jiang M, Xiao R, Soong T, Rost B, Acton TB, Everett JK, Pegg AE, Tainer JA, Montelione GT (2010) Solution NMR structure of a bacterial alkyltransferase-like protein. **J. Biol. Chem** (Submitted).
24. Mao L, Tao Y, Inouye K, Montelione GT, McDermott AE, Inouye M (2010) Solid state NMR spectra of a membrane protein in a natural membrane environment. **J. Am. Chem. Soc.** (Submitted).

25. Price WN, Handelman SK, Everett J, Tong SN, Bracic A, Acton T, Xiao R, Rost B, Montelione GT, Hunt JF (2010) Large scale studies show unexpected amino acid effects on protein expression and solubility. **Protein Science** (Submitted).
26. Nie Y, Xiao R, Xu Y, Montelione GT (2010) Identification and characterization of cluster of a novel anti-prelog type stereospecific carbonyl reductase in the genome of *Candida parapsilosis*. **ACS Chemical Biology**. (Submitted).
27. Zhao L, Zhao K, Hurst R, Slater M, Acton TB, Swapna GVT, Shastri R, Kornhaber GJ, Montelione GT (2010) Engineering of a wheat germ expression system to provide compatibility with a high throughput pET-based cloning platform. **J. Struct. Funct. Genomics** (Submitted).
28. Arbing MA, Handelman SK, Kuzin AP, Verdon G, Su M, Abashidze M, Liu M, Hurley JM, Xiao R, Acton T, Inouye M, Montelione GT, Woychik NA, Hunt JF (2010) Structural similarity establishes multiple evolutionary links between microbial growth-regulating toxin and antitoxin (TA) systems. **J. Mol Biol.** (Submitted).
29. Lee H-W, Wylie G, Bonsal S, Wang X, Barb AW, Macnaughtan MA, Ertekin A, Montelione GT, Prestegard JH (2010) Determination of three-dimensional structures for weakly associated protein homodimers using RDCs and paramagnetic surface mapping. **J. Am. Chem. Soc.** (Submitted).
30. Zhang, H, Constantine, R, Vorobiev, S, Chen, Y, Seetharaman, J, Huang, YJ, Xiao, R, Montelione, GT, Davis, MW, Inana, G, Jorgensen, EM, Tong, L, Baehr, W (2010) UNC119 regulates G protein trafficking in sensory neurons by recognizing the acylated G α N-terminus. **Cell** (Submitted).
31. Norel R, Petrey D, Honig B (2010) PUDGE: A flexible, interactive server for protein structure prediction. **Nucleic Acid Research** (Submitted).
32. Mercier KA, Stark JL, Mueller GA, Acton TB, Xiao R, Montelione GT, Powers R (2010) Solution structure and function of YndB, an AHS1 protein from *Bacillus subtilis*. **PROTEINS: Struct. Funct. Bioinformatics** (Submitted).
33. Tang Y, Xiao R, Ciccocanti, C, Janjua H, Lee YL, Everett J, Swapna GVT, Acton TB, Rost B, Montelione GT. (2010) Solution NMR structure of Lin0431 protein from *Listeria innocua* reveals high structural similarity with domain II of bacterial transcription antitermination protein NusG. **PROTEINS: Struct. Funct. Bioinformatics** (Submitted).
34. Handelman SK, Tong SN, Luff JD, Lee DP, Lazar A, Smith P, Shao Y, Conway T, Hunt JF (2010) Classes of reciprocal sequence homologues (CRSHs): orthology-based groups of proteins of likely equivalent function for analysis of bacterial protein physiology and evolution. **J. Mol. Biol.** (Submitted)
35. Arragain S, Handelman SK, Forouhar F, Fan-Yan W, Douki T, Fontecave, Hunt JF, Mulliez ME, Atta, M. (2010) Experimental identification of eukaryotic and prokaryotic methylothiotransferases for biosynthesis of 2-methylthio-N⁶-threonylcarbamoyladenine in tRNA. **J. Biol. Chem.** (Submitted)