Ram Mani
June 2, 2000

Update for COG Computational Work - COG's 316S & 229S

COG 229S

This COG continues to have **no structural homologues**. It also appears **that no competing scientists** have indicated that they are working on structural determination of this protein domain. Several sequences of the COG were searched against sites listed on the CABM NMRLab Structural Genomics "Other Structural Genomics Projects" web page[1]. No similar sequences were found. The 229 *C. elegans* sequence was searched against the PDB sequence database using PSI-BLAST[2]. No similar sequences were found.

There is some functional significance for this domain. The 229 domain is found in some multi-domain proteins exhibiting a peptide methionine sulfoxide reductase (PMSR) activity. However, the domain that actually performs this reductase function is a domain other than the 229 domain (PMSR activity domain has InterPro Entry IPR002569)[3]. These PMSR proteins repair damaged proteins.

The 229 domain is present in bacterial proteins of the pilB family of transcriptional factors (pilin). These proteins are involved in regulation of genes that code for components of the bacterial pilus. The literature describing focuses on the pilB *N. gonorrhoeae* multi-domain protein containing the 229 domain[4]. It is suggested that this protein may interact with other(s) to regulate transcription. A Yale group has also recently investigated an *H. sapiens*, hypothetical, 182 residue single-domain, gene product that contains the 229 domain[5]. Their preliminary studies shows that the protein is expressed ubiquitously in many tissues, but most abundantly in the retina and ocular ciliary body, skeletal muscle, and heart. They suggest the gene (CBS-1) for this protein encodes a mammalian transcription factor.

The InterPro entry for this protein (InterPro code IPR002579) suggests that this domain of unknown function may contain a zinc-binding site because of its two conserved cysteines and histidines. This InterPro entry also reports that the final cysteine is found to be replaced by selenocysteine in some members of this domain family[6].

To pursue in the future

A new multiple sequence alignment will be done (Greg Kornhaber, *et al.*) to includes the full-length *C. elegans* sequence that I have begun characterizing and new sequences. I suspect that the N-terminus of the sequence that I am expressing may contain several non-conserved residues of the domain. They may contribute to the unfolded region of this protein (see thesis). In addition, limited proteolysis may cleave away this disordered region and result in a stable core. Another route to take is addition of metal ions, such as zinc, to the protein solution. SDS-PAGE/CD/NMR will be used to screen for a stable protein.

---

[1] http://www-nmr.cabm.rutgers.edu/structuralgenomics/other.html
Sites consulted were: (1) Fold Diversity Pilot Project of Rockefeller, Einstein, and BNL (S. Cerevisiae); (2) Presage Structural Genomics Database; (3) PSI Target Database; (4) Structure to Function Pilot Project of CARB & TIGR (H. Influenza)

[2] http://www.ncbi.nlm.nih.gov/blast/psiblast.cgi

[3] InterPro web site is http://interpro.ebi.ac.uk

[4] Taha, M.K., So, M., Seifert, H.S., Billyard, E., and Marchal, C. (1988) Pilin expression in Neisseria gonorrhoeae is under positive and negative transcriptional control. *EMBO J.* 7, 13: 4367-78.

[5] Huang, W., Escribano, J., Safarazi, M., and Coca-Prados, M. (1999) Identification, expression and chromosome localization of a human gene encoding a novel protein with similarity to the pilB family of transcriptional factors (pilin) and to bacterial peptide methionine sulfoxide reductases. *Gene*. 233, 1-2: 233-40.

[6] Lescure, A., Gautheret, D., Carbon, P., and Krol, A. (1999) Novel selenoproteins identified in silico and in vivo by using a conserved RNA structural motif. *J. Biol. Chem.* 274, 53: 38147-54.

Large quantities of the protein (7.5 to 10 mg from 250 ml of culture) have been isolated by expressing at 17°C and purifying using nickel-affinity chromatography (see figure 1; I want your estimate on the purity of this sample). I have been performing gel filtration on the sample in the last few days, and this step does not look promising because the protein is not eluting from the column. Gel filtration has not been rigorously tested because problems in concentrating the protein using an amicon allow only for loading of a 0.5 mg/ml solution (1 ml total) onto the column. Because of these problems with gel filtration, I may go directly from a nickel-affinity chromatography purified sample to limited proteolysis.

COG 316

This COG has **no structural homologues** detectable from sequence searching using the 316 *H. Sapiens*-2 sequence[2]. The CARB/TIGR Structure 2 Function *H. Influenza* web site reports the *H. Influenza*-1 sequence is in the **NMR data collection step**[7]. This report of NMR data collection was indicated on the web site sometime between May 18 and May 26, 2000.

The CARB/TIGR web site also has a submitted prediction from David Jones stating that the structure for this sequence may be similar to verotoxin-1 (PDB ID 1bovA), which is also known as Shiga-like toxin-1. The method used for this structure prediction is GeneThreadFold. When comparing the 1bovA sequence to the 316 *H. Influenza*-1 sequence, a BLAST2 comparison gives a E value significance score of 3654, 26% identity, and 42% similarity over a region of 61 residues. Based on these data and inspection for the highly conserved residues of the 316 domain family, there is no detectable homology between the 1bovA and the 316 *H. Influenza*-1 sequence.

The Rockefeller/Einstein/BNL web site indicates that this group has selected the 316 *S. Cerevisae*-1 as a target (project code P114) for expression and characterization, but they have not indicated any progress towards the 316 sequence and they have not updated their web site since September 1999.

My reagents used for expressing this domain have been placed in long-term storage and are being entered in the lab's electronic, reagent database.

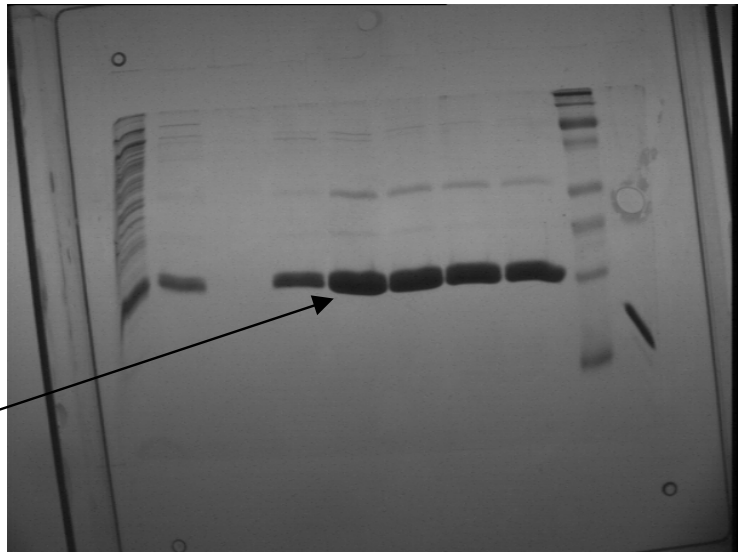

Figure 1

[7] http://s2f.carb.nist.gov/cgi-bin/display.cgi?geneid=HI1723