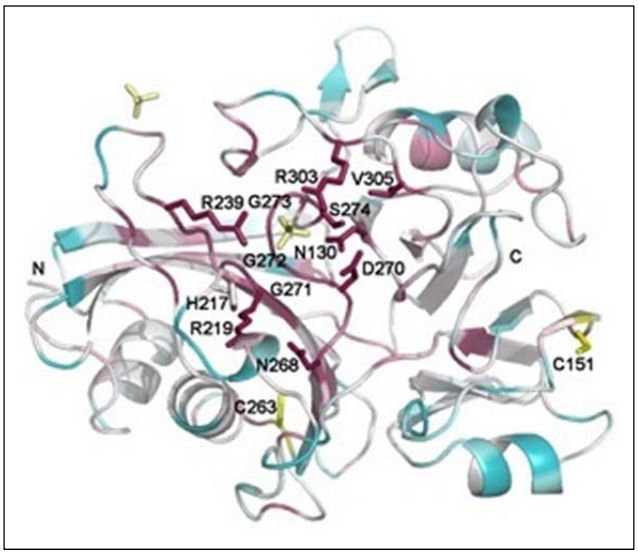
Structure of the DUF2233 Domain in Bacteria and the Stuttering-associated UCE Glycoprotein

DUF2233, a Domain of Unknown Function (DUF), is present in ~1200 bacterial and several viral and eukaryotic proteins. DUF2233 has been identified in proteins ranging in size from ~300-2000 residues. The 515 amino acid mammalian transmembrane glycoprotein α-N-acetylglucosamine-1-phosphodiester N-acetylglucosaminidase ("Uncovering enzyme", UCE), which is localized in the *trans*-Golgi network, is the only human protein that contains the DUF2233 domain.



UCE is important in a cellular recycling system. UCE converts N-acetylglucosamine-P-mannose diester to mannose-6-P monoester on newly synthesized lysosomal acid hydrolases, a key step in the targeting of these hydrolases to lysosomes. Disruption of the *Nagpa* gene that codes for UCE leads to excessive cellular secretion of acid hydrolases. Recently, *Nagpa* mutations have been associated with persistent stuttering in humans. Despite the importance of UCE, very limited information is available concerning its structure. Nothing is known about the function of the prokaryotic DUF2233 members.

Using the JCSG High-Throughput Structural Biology platform and SSRL Beam Line 9-2, the first crystal structure of a representative of the DUF2233 protein family, BACOVA_00430 from *Bacteroides ovatus* a human gut bacteria, has now been solved at a resolution of 1.80 Å. BACOVA_00430 consist of 4 domains, each of which resembles to some extent the cystatin fold, which consists of a curved anti-parallel β-sheet wrapped around an α-helix. The first domain, which is not included in the DUF2233 definition, resembles more closely the typical cystatin-like fold. Domains 3 and 4 can be superimposed on each other (r.m.s.d. of 2.2 Å over 41 Cα atoms and 22% sequence identity), suggesting possible gene duplication in this portion of the protein. A search for other proteins with similar structure did not produce any significant results, thereby revealing that BACOVA_00430 is the first representative of the novel DUF2233 family. Comparative sequence analysis of bacterial DUF2233 members revealed several conserved residues located in a cleft on the surface of the protein, indicating that they are likely involved in function. Amongst these, Asn268, Asp270, Gly271-273 and Ser274 are part of the highly conserved A(I/L)NLDGGGS(T/S/A)T motif present throughout the DUF2233 family (Figure, in pink sticks). A sulfate ion from the crystallization condition is bound near Gly272-Gly273 of the GGGS sequence, anchored by conserved residues Arg239 and Arg303.

The BACOVA_00430 structure was used as a template for modeling the luminal region of UCE. Model-based site-directed mutagenesis of UCE in Dr. Stuart Kornfeld's laboratory at Washington University School of Medicine in St. Louis confirmed the predicted functional importance of most of these conserved residues as judged by comparison with the activity of wild-type UCE and its influence on trafficking to the Golgi. Similar mutational analyses were performed on BACOVA_00430. Kinetics studies of the UCE and BACOVA_00430 on different substrates were also performed. These indicated that both proteins function as phosphodiester glycosidases and UCE has evolved to specifically recognize and act on

GlcNAc-P-Mannose. To date, the physiological substrate of BACOVA_00430 or other bacterial DUF2233 proteins remains unknown. However, based on the assignment of function as a phosphodiester glycosidase, a number of bacterial cell wall components with sugar-P-sugar repeating structures could potentially be substrates for BACOVA_00430 and related bacterial proteins. The sulfate from the crystallization condition found near Gly272-Gly273 and Arg239/Arg303 may mimic a phosphate from the physiologically relevant substrate. Residues Asn130, Asp270 and Ser274 are potential catalytic residues. These studies provide the first structure-function analysis of DUF2233 proteins.

Primary Citation

D. Das, W.-S. Lee, J. C. Grant, H.-J. Chiu, C. L. Farr, J. Vance, H. E. Klock, M. W. Knuth, M. D. Miller, M.-A. Elsliger, A. M. Deacon, A. Godzik, S. A. Lesley, S. Kornfeld and I. A. Wilson, "Structure and Function of the DUF2233 Domain in Bacteria and in the Human Mannose 6-Phosphate Uncovering Enzyme", *J. Biol. Chem.* **288**, 16789 (2013) doi: 10.1074/jbc.M112.434977

Contact

Debanu Das, JCSG, SSRL Structural Genomics

SSRL is primarily supported by the DOE Offices of Basic Energy Sciences and Biological and Environmental Research, with additional support from the National Institutes of Health, National Institute of General Medical Sciences.