NMR Structure of the 18 kDa Protein CC1736 From Caulobacter crescentus Identifies a Member of the “START” Domain Superfamily and Suggests Residues Mediating Substrate Specificity

Yang Shen,1,4 Sharon Goldsmith-Fischman,2,4 Hanudatta S. Atreya,1,4 Thomas Acton,3,4 LiChung Ma,3,4 Rong Xiao,3,4 Barry Honig,2,4 Gaetano T. Montelione,3,4 and Thomas Szyperski1,4*

1Department of Chemistry, University at Buffalo, the State University of New York, Buffalo, New York
2Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York
3Center of Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey
4Northeast Structural Genomics Consortium

Introduction. The 18-kDa protein CC1736 (SwissProt/TrEMBL ID Q9A7I7) of Caulobacter crescentus belongs to a target cluster1 of the Northeast Structural Genomics Consortium (NESGC; http://www.nesg.org; NESGC target ID: CcR19; NESGC Rost-cluster ID: 17538) comprising both eukaryotic and bacterial proteins. CC1736 is member of Pfam2 “Aromatic-rich protein family” (Pf03654) and has at least 90 close sequence homologs. CC1736 exhibits weak sequence homology with oligoketide cyclases and aromata- ses which bind multi-cyclic and/or aromatic compounds such as cholesterol (or polyketides).3

Materials and Methods. As described elsewhere,4 resonance assignments were obtained from NMR data collected for a 13C,15N-labeled CC1736 sample (1.1 mM) at 25°C using a Varian INOVA 750 spectrometer. Upper-distance limit constraints were obtained from 3D15N- and 13C-resolved [1H,1H]-NOESY5 (Table I), 3JHN scalar couplings were measured in 3D HNNHA 5 yielding ϕ-angle constraints, and backbone dihedral angle constraints were derived from chemical shifts as described6 for residues in secondary structure elements. Structure calculations were performed using the program DYANA.7

Results and Discussion. The NMR structure {Table I, Fig. 1(a)} of CC1736 (PDB ID: 1T17) reveals two α-helices I and II and seven β-strands A–F {Fig. 1(b)} giving rise to an α+β fold in which helices and strands are segregated. The β-strands are arranged with topology A(†), G(⊥), F(†), E(⊥), D(†), C(⊥), B(†) and form a highly twisted β-sheet. The juxtaposition of α-helices and β-strands leads to the formation of a hydrophobic tunnel, which is likely of functional importance (see below).

A search for structurally similar domains using the programs DALI,8 SKAN,9 and CE10 reveals that CC1736 is structurally similar to birch pollen allergen (PDB ID: IBV1, DALI Z-score 12.0), phosphatidylinositol transfer protein (PDB ID: IFVZ, DALI Z-score 8.1), phosphoglucomutase domain (PDB ID: 3PMG, DALI Z-score 6.0), and the cholesterol-regulated START domain (PDB ID: LJSS, DALI Z-score 11.5). According to SCOP11 classification, these proteins exhibit a “TBP-like” (TATA-Binding Protein-like) fold and they belong to the START Domain Superfamily,12 which includes oligoketide cyclases.

Sequence comparisons show that CC1736 is: i) a member of the “Aromatic-Rich Protein Family” (Pfam2 ID: PF03654), which is annotated as having a possible relationship to polyketide synthases/cyclases, and ii) exhibits modest sequence homology with oligoketide cyclases such as the WhiE putative polyketide cyclase (26% identity, BLAST E-value = 0.17, member of “Polyketide Cyclase Family” Pfam ID: PF03654). No sequence similarity is detected between CC1736 and START domain even in iterative PSI-BLAST13 searches. However, the fold recognition program HMAP14 recognizes a relationship between CC1736 and START domain,8 supporting without reference to its NMR structure that CC1736 is a member of the START domain superfamily.