Functional Implications from Crystal Structures of the Conserved *Bacillus subtilis* Protein Maf with and without dUTP

G. Minasov,[1] M. Teplova,[2] G. C. Stewart,[3] E. V. Koonin,[4] W. F. Anderson,[1] M. Egli [2,*]

[1] *Department of Molecular Pharmacology and Biological Chemistry, Northwestern University Medical School, Chicago, IL, U.S.A.*

[2] *Department of Biological Sciences, Vanderbilt University, Nashville, TN, U.S.A.*

[3] *Department of Diagnostic Medicine/Pathobiology, College of Veterinary Medicine, Kansas State University, Manhattan, KS, U.S.A.*

[4] *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, U.S.A.*

**To whom correspondence should be addressed: martin.egli@ vanderbilt.edu)*

Abstract

Three-dimensional structures of functionally uncharacterized proteins may furnish insight into their functions. The potential benefits of 3D-structural information regarding such proteins are particularly obvious when the corresponding genes are conserved during evolution and no functional classification can be inferred from their sequences. The current accumulation of vast amounts of sequence information and the identification of conserved open reading frames (ORFs) of unknown function creates an urgent need to exploit novel strategies to investigate function. Wide phylogenetic distribution of such ORFs may imply that the encoded protein serves an important function. The *Bacillus subtilis* Maf protein is representative of a family of proteins that has homologues in many of the completely sequenced genomes from archaea, prokaryotes, and eukaryotes, but whose function is unknown. As an aid in exploring its function, we determined its crystal structure at a resolution of 1.85 Å. The structure, in combination with multiple sequence alignment, reveals a putative active site. Phosphate ions present at this site and structural similarities between a portion of Maf and the anticodon-binding domains of several tRNA synthetases suggest that Maf may be a nucleic acid-binding protein. The crystal structure of a Maf-nucleoside triphosphate complex provides support for this hypothesis and hints at di- or oligonucleotides with either 5′- or 3′-terminal phosphate groups as ligands or substrates of Maf. A further clue comes from the observation that the structure of the Maf monomer bears similarities to that of the recently reported *Methanococcus jannaschii* Mj0226 protein. Just as for Maf, the structure of this predicted NTPase was determined as part of a structural genomics pilot project. The structural relation between Maf and Mj0226 was not apparent from sequence analysis approaches. These results emphasize the potential of structural genomics to reveal new, unexpected connections between protein families previously considered unrelated.

Methods and Materials

Crystals suitable for data collection were obtained within one day with both wt and Se-Met Maf. All crystals were shock-frozen in the mother liquor plus 25% sucrose. Diffraction data for wt Maf were collected at 100 K on the 5-ID beamline of the DND-CAT at the APS to a maximum resolution of 1.85 Å. MAD data were collected at four wavelengths on a single Se-Met Maf crystal to a resolution of 2.5 Å. All data were integrated and scaled in the DENZO/SCALEPACK suite. Se sites were determined with the program SOLVE based on 2.8 Å anomalous data, and 8 selenium atoms per asymmetric unit could be located. The asymmetric unit consists of two Maf molecules, each containing 6 Se atoms. Se sites were refined with the program SHARP, but no new Se sites were found. Electron density maps were calculated in the CCP4 suite, and about 85% of the polypeptide backbones of both molecules were readily built using the program O. Cycles of manual rebuilding were followed by positional, simulated annealing, and temperature factor refinements with the program CNS and gradually improved the model. Along with two Maf molecules, 304 water molecules, three phosphate ions, and one sucrose molecule were built into the electron density maps. The final R-factor is 19.5% (R-free 22.3%) for 40,372 reflections in the 25.0-1.85 Å resolution range (bulk solvent correction).

Acknowledgments

This work was supported by National Institutes of Health grant GM57049 (G.C.S.) and partially supported by NIH grant GM55237 (M.E.). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-ENG-38. The DuPont-Northwestern-Dow Collaborative Access Team Synchrotron Research Center, located at Sector 5 of the Advanced Photon Source, is supported by E. I. DuPont de Nemours & Co., The Dow Chemical Company, the National Science Foundation, and the State of Illinois.

Reference

G. Minasov, M. Teplova, G. C. Stewart, E. V. Koonin, W. F. Anderson, and M. Egli, Proc. Natl. Acad. Sci. U.S.A. **97**, 6328-6333 (2000).