doi:10.1016/j.jmb.2008.03.007 J. Mol. Biol. (2008) 379, 284–298

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Structure and Function of the Arginine Repressor-Operator Complex from *Bacillus subtilis*

James A. Garnett, Ferenc Marincs, Simon Baumberg†, Peter G. Stockley and Simon E. V. Phillips*

Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

Received 17 December 2007; received in revised form 29 February 2008; accepted 3 March 2008
Available online 12 March 2008

In many bacteria, the concentration of L-arginine is controlled by a transcriptional regulator, the arginine repressor. In *Bacillus subtilis* this transcription factor is called AhrC and has roles in both the repression and activation of the genes involved in arginine metabolism. It interacts with 18 bp ARG boxes in the promoters of arginine biosynthetic and catabolic operons. AhrC is a hexamer and each subunit has two domains. The C-terminal domains form the core, mediating inter-subunit interactions and L-arginine binding, while the N-terminal domains contain a winged helix-turn-helix DNA-binding motif and are arranged around the periphery. Upon binding of the co-repressor L-arginine there is a ~15° relative rotation between core C-terminal trimers. Here, we report the X-ray crystal structure of a dimer of the N-terminal domains of AhrC (NAhrC) in complex with an 18 bp DNA ARG box operator, refined to 2.85 Å resolution. Comparison of the N-terminal domains within this complex with those of the free domain reveals that the flexible β-wings of the DNA-binding motif in the free domain form a stable dimer interface in the protein–DNA complex, favouring correct orientation of the recognition helices. These are then positioned to insert into adjacent turns of the major groove of the ARG box, whilst the wings contact the minor groove. There are extensive contacts between the protein and the DNA phosphodiester backbone, as well as a number of direct hydrogen bonds between conserved amino acid side chains and bases. Combining this structure with other crystal structures of other AhrC components, we have constructed a model of the repression complex of AhrC at the *B. subtilis* biosynthetic *argC* operator and, along with transcriptome data, analysed the origins of sequence specificity and arginine activation.

Keywords: arginine repressor/activator; crystal structure; structure-function; gene array; transcriptional control

Edited by K. Morikawa

Introduction

The genetic control of arginine-metabolising enzymes in many bacteria is achieved through the arginine repressor,[1–4] which responds to intracellular levels of L-arginine. In *Escherichia coli*, where the arginine repressor (ArgR) was first identified, the protein controls a regulon of mostly separate genes involved in arginine biosynthesis.[3,5] It is also an essential component in the protein–DNA complex that resolves plasmid multimers during their replication.[6] Bioinformatic analysis has revealed putative homologues of ArgR in many species.[7–9]

The L-arginine analogue, L-arginine hydroxamate, is an antagonist of L-arginine, and *Bacillus subtilis* mutants resistant to it overproduce L-arginine.[10] Har-

*Corresponding author. E-mail address: s.e.v.phillips@leeds.ac.uk.

Present addresses: J. A. Garnett, Division of Molecular Biosciences, Imperial College London, South Kensington, London SW7 2AZ, UK; F. Marincs, Agricultural Biotechnology Center, Microarray Laboratory, Gödöllö, Hungary.

† Deceased 11th April 2007.

Abbreviations used: ArgR, arginine repressor; AhrC, arginine hydroxamate resistant mutant C; NCS, non-crystallographic symmetry.

wood and Baumberg were able to map the resistance to four loci.[11] One of these was identified as a homologue of *E. coli argR*, and it was termed the arginine hydroxamate resistant mutant C (AhrC). The AhrC protein shares 27% overall amino acid sequence identity with *E. coli* ArgR, which rises to 35% in the C-terminal domain.[4] In *B. subtilis*, in addition to acting as a repressor of arginine biosynthetic genes, AhrC functions as an activator of arginine catabolic genes. Both the biosynthetic and catabolic genes are clustered into two operons that start with *arg*C and *argG*,[12–14] and *rocA* and *rocD*,[13,15,16] respectively.

AhrC and *E. coli* ArgR have been used as the primary models in which the basic features of these

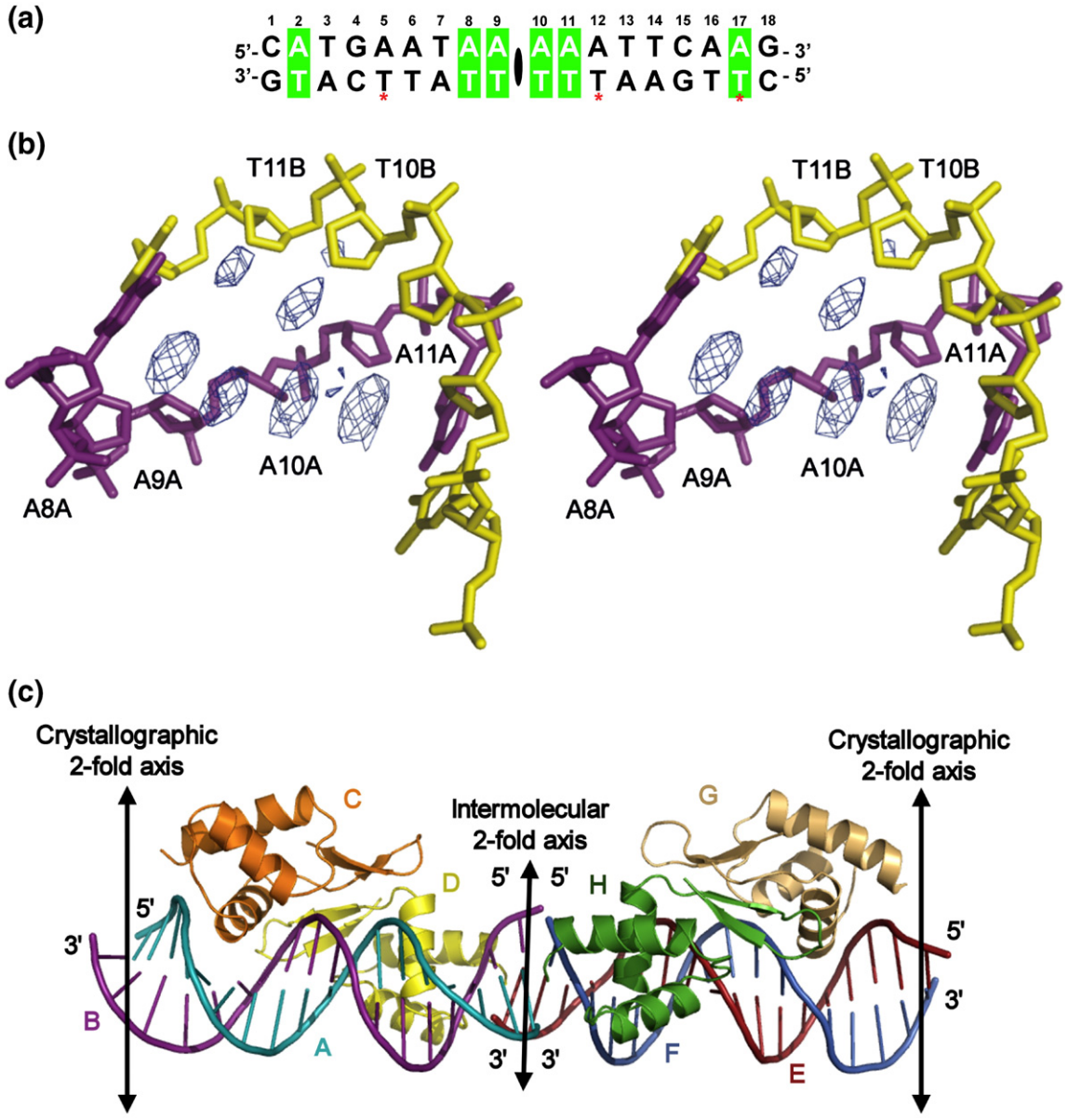

Fig. 1. Structure and symmetry of the N-terminal AhrC-DNA complex in the asymmetric unit. (a) ARG box sequence used in the NAhrC–DNA complex crystal structure. The top strand corresponds to polynucleotide chains A and E of the complex, and the bottom strand to polynucleotide chains B and F. An approximate dyad passes between base pairs 9 and 10. Asymmetric base positions are shaded green and positions that were substituted by 5-bromouracil in the Br-NAhrC–DNA complex are identified by a red asterisk. (b) Electron density showing the orientation of the ARG box DNA in the asymmetric unit for chains A (purple) and B (yellow). Asymmetric bases were removed and the $|F_o|-|F_c|$ map calculated after refinement (contoured at 5.5 σ). The larger peaks correspond to adenine and are associated with chain A, while the smaller peaks (thymine) belong to chain B. (c) Relationship between the two independent complexes in the asymmetric unit, in a similar orientation to that in (b). ARG box chains A (teal), B (purple), E (red), F (blue) and NAhrC chains C (orange), D (yellow), G (peach), H (green) pack in the asymmetric unit, with chains A–D and chains E–H related by approximate intermolecular twofold NCS. The adjacent complexes in the lattice are stacked end to end, related by twofold crystallographic symmetry to form pseudo-continuous duplexes. (d) The triple base pair formed at the NCS intersection, where two DNA duplexes pack 5′ to 5′, and 3′ to 3′. (e) The structure of the NAhrC-DNA complex, chains A –D. The two half-complexes are related by intramolecular pseudo twofold NCS.