doi:10.1016/j.jmb.2008.10.041 J. Mol. Biol. (2009) 385, 469–490

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Crystallographic Analysis of a Sex-Specific Enhancer Element: Sequence-Dependent DNA Structure, Hydration, and Dynamics

Narendra Narayana* and Michael A. Weiss

Department of Biochemistry, Case Western Reserve University, Cleveland, OH 44106, USA

Received 5 June 2008; received in revised form 2 October 2008; accepted 6 October 2008
Available online 22 October 2008

The crystal structure of a sex-specific enhancer element is described at a resolution of 1.6 Å. This 16-bp site, designated Dsx^A, functions in the regulation of a genetic switch between male and female patterns of gene expression in *Drosophila melanogaster*. Related sites are broadly conserved in metazoans, including in the human genome. This enhancer element is unusually rich in general regulatory sequences related to DNA recognition by multiple classes of eukaryotic transcription factors, including the DM motifs, homeodomain, and high mobility group box. Whereas free DNA is often crystallized as an A-form double helix, Dsx^A was crystallized as B-DNA and thus provides a model for the prebound conformation of diverse regulatory DNA complexes. Sequence-dependent conformational properties that extend features of shorter B-DNA fragments with respect to double helical parameters, groove widths, hydration, and binding of divalent metal ions are observed. The structure also exhibits a sequence-dependent pattern of isotropic thermal *B*-factors, suggesting possible variation in the local flexibility of the DNA backbone. Such fluctuations are in accord with structural variability observed in prior B-DNA structures. We speculate that sites of intrinsic flexibility within a DNA control element provide hinges for its protein-directed reorganization in a transcriptional preinitiation complex.

Keywords: DM domain; DNA structure; Doublesex; protein–DNA recognition; sex determination

Edited by B. Honig

Introduction

DNA structure is fundamental to diverse DNA transactions, such as replication, repair, nucleosome assembly, and the regulation of gene expression.[1,2] Architectural and regulatory proteins bind to DNA with an affinity, specificity, and kinetics appropriate to their biological roles.[3] Comparative structural analyses of DNA oligonucleotides and protein–DNA complexes have provided insight into the physical principles of DNA structure, reorganization, and recognition. DNA–protein recognition is enabled by the characteristic structure of the double helix, including base-specific arrays of functional groups in the major and minor grooves,[4] specific water molecules,[5] and counterion distributions.[6,7] Of particular interest is the potential contribution of DNA dynamics and induced fit[8–11] to recognition: should specific B-DNA sequences differ in malleability,[12,13] such rigid or flexible structural elements may contribute to "indirect readout" of the underlying DNA sequence.[5,14,15] In this article, we describe at high resolution the crystal structure of an enhancer DNA element that functions in a well-characterized genetic switch, the reciprocal regulation of male and female gene expression by the Doublesex (DSX) proteins in *Drosophila melanogaster*.[16] Designated dsx^A, this enhancer element contains subsites related to multiple classes of

*Corresponding author. E-mail address: narayanan3@uthscsa.edu.

Present address: N. Narayana, Department of Biochemistry, The University of Texas, Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78229-3900, USA.

Abbreviations used: BrU, 5-bromo-deoxyuridine; DDD, Dickerson–Drew dodecamer; DM, N-terminal DNA-binding domain of DSX; DSX, Doublesex; Exd, Extradenticle; HMG, high mobility group; MAD, multiple anomalous dispersion; MIS, Mullerian inhibiting substance; PEG, polyethylene glycol; Scr, Sex combs reduced.

eukaryotic transcription factors (Table 1). Because in most cases structures of protein–DNA complexes have been determined in the absence of the structures of their cognate free DNA sites, the structure of *dsx*^{A} provides a model for general analysis of sequence-specific recognition and induced fit by diverse DNA-binding motifs. Our results extend previous crystallographic studies of sequence-dependent structural features in shorter oligonucleotides and provide novel evidence for the sequence-dependent flexibility of B-DNA at discrete sites of local backbone disorder.

Because of their fundamental importance, crystallographic studies of DNA fragments have been pursued for more than two decades. Much of our understanding of sequence-dependent features of free B-DNA structures stems from extensive structural studies of the Dickerson–Drew dodecamer [DDD; d(CGCGAATTCGCG)$_2$] and related sequence isomers and decamers[12] (Throughout the text, references to DDD or R3 dodecamer crystal form relate to the sequence d(CGCGAATTCGCG)$_2$.) These structures exhibit sequence-dependent conformations with respect to base-pair geometry, backbone flexibility, groove widths and depths, hydration, metal ion coordination, bending, and deformability.[12,17] Two key features are variation of the width of the minor groove and bending of the DNA double helix. Such conformational variations contribute to sequence-specific recognition of DNA by proteins. In addition to geometric features of the double helix, the binding of water molecules and cations at specific sites is influenced by the functional groups of DNA bases.[6] Although the positions and types of cations surrounding the DNA duplex can be important determinants of lattice packing and influence local DNA conformation,[7] the broad relevance of crystallographic features to the structure of DNA in solution is well supported by NMR studies based on nuclear Overhauser effects[18] and residual dipolar couplings.[19]

Table 1. Cognate DNA subsites (5′→3′) of transcription factors found in the present structure

Protein	DNA target site
A. *The DSX and the shared SRY binding sites*	
Present study (16-bp duplex)	5′-ACTACAATGTTGCAAT-3′
DSX[a]	ACAATGTT
SRY[b]	ACAAT
B. *Subsites (four or more base pairs) found within the 16-bp duplex*	
FOXJ2	ACAAT
Sox-5/Sox-9[c]	ACAAT
Mat1-Mc	ACAAT
HNF-3β[d]	AATGTT
C/EBP[e]	AATGTTGCAAT
CHOP-C/EBPα[f]	TGCAAT
FOXD3	ATGTT
Oct-1[g]	CAATGT
MIF-1	GTTGC
VBP	TACA
RFX1	GTTGCAAT
HLF	GCAAT
MATa1[h]	ATGT
Elf-1	CTACA
Athb-1	TTGCA
AbaA	AATGT
c-Myb[i]	GTTG
C. *Subsites spanning the pseudo-continuous helical section of the duplexes in the crystal*	
v-Myb	AATA (…GC*AAT*-3′ adjoining 5′-*A*CTC…)
GATA-1[j]	ATACT (…GCA*AT*-3′ adjoining 5′-*ACT*AC…)
CDP CR3	TATTG (…C*GTTA*-5′ adjoining 3′-*T*GAT…)

SRY, sex-determining region Y protein; FOXJ2, fork head box J 2; Sox, SRY-related HMG-box gene; Mat1-Mc, M-box interacting with Mat1-Mc; HNF-3β, hepatocyte nuclear factor 3β; C/EBP, C/EBP binding site; CHOP-C/EBPα, heterodimers of CHOP and C/EBPα; FOXD3, fork head box D3; Oct-1, octamer-binding factor 1; MIF-1, MIBP-1/RFX1 complex; VBP, PAR-type chicken vitellogenin promoter-binding protein; RFX1, X-box binding protein RFX1; HLF, hepatic leukemia factor; MATa1, mating factor a1; Elf-1, Elf-1 binding site; Athb-1, *Arabidopsis thaliana* homeo box protein 1; AbaA, AbaA binding site; c-Myb, c-Myb binding site; v-Myb, viral myb binding site; GATA-1, GATA-binding factor 1; CDP CR3, cut-like homeodomain protein. The availability of NMR or crystal structures (protein or its homolog or DNA target site or protein–DNA complex) of a specific transcription factor is indicated by a superscript letter, and the respective Protein Data Bank codes are as follows: (a) 1lpv, (b) 1lwa and 1j46, (c) 1i11, (d) 2hfh, (e) 2e43, (f) 1nwq, (g) 1hf0, (h) 1f43, (i) 1bdz and 1idy, and (j) 3gat.

We describe herein the crystal structure of an A+T-rich DNA duplex (5′-ACACTACAATGTTGC-AAT-3′ and complement; Fig. 1, top panel) derived from the *Drosophila* fat body enhancer.[20] The central segment (underlined) defines *dsx*^{A}, a control site that contributes to the sex- and tissue-specific expression of yolk proteins by the male and female isoforms of the DSX transcription factor.[20] This structure (16 bp; Fig. 1, bottom panel) represents the longest free B-DNA duplex structure to be characterized at atomic resolution (1.6 Å) to date, sufficient to visualize sequence-dependent variations in structure, hydration, and metal-ion binding. The sequence is notable for a central A+T-rich segment (5′-AATGTT-3′) critical to recognition by the conserved DM family of minor-groove DNA-binding proteins.[21,22] Although the DNA site is not self-complementary, the structure exhibits a striking pseudo-2-fold symmetry with respect to base-pair and sugar-phosphate geometries, groove widths, hydration patterns, and metal-cation distribution. These parameters appear to be context specific as well as sequence specific. The central portion of the *dsx*^{A} site (5′-AATGTT-3′) is straight with B-type helical parameters consistent with previous A+T-rich B-DNA structures;[23] the adjoining three base-pair steps on either side resemble an alternating B-type conformation.[24] The two terminal base-pair steps at each end are slightly bent relative to the central site. Because subsites within *dsx*^{A} resemble target sites for diverse classes of eukaryotic transcription factors contacting either or both DNA grooves (Table 1), their structural features within free DNA may be of broad biological interest.

An unexpected feature of *dsx*^{A} relates to its pattern of crystallographic order and disorder. Whereas variations in mean atomic coordinate position are consistent with the database of DNA structures,[23] thermal *B*-factors exhibit anomalies. Specific deoxy-