doi:10.1016/j.jmb.2009.06.062 J. Mol. Biol. (2009) 392, 115–128

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Crystal Structure of the HEAT Domain from the Pre-mRNA Processing Factor Symplekin

Sarah A. Kennedy[1], Monica L. Frazier[2], Mindy Steiniger[3], Ann M. Mast[1], William F. Marzluff[2,3] and Matthew R. Redinbo[1,2*]

[1]*Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA*

[2]*Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA*

[3]*Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA*

Received 27 March 2009; received in revised form 15 June 2009; accepted 25 June 2009
Available online 1 July 2009

The majority of eukaryotic pre-mRNAs are processed by 3′-end cleavage and polyadenylation, although in metazoa the replication-dependent histone mRNAs are processed by 3′-end cleavage but not polyadenylation. The macromolecular complex responsible for processing both canonical and histone pre-mRNAs contains the ~1160-residue protein Symplekin. Secondary-structural prediction algorithms identified putative HEAT domains in the 300 N-terminal residues of all Symplekins of known sequence. The structure and dynamics of this domain were investigated to begin elucidating the role Symplekin plays in mRNA maturation. The crystal structure of the *Drosophila melanogaster* Symplekin HEAT domain was determined to 2.4 Å resolution with single-wavelength anomalous dispersion phasing methods. The structure exhibits five canonical HEAT repeats along with an extended 31-amino-acid loop (loop 8) between the fourth and fifth repeat that is conserved within closely related Symplekin sequences. Molecular dynamics simulations of this domain show that the presence of loop 8 dampens correlated and anticorrelated motion in the HEAT domain, therefore providing a neutral surface for potential protein–protein interactions. HEAT domains are often employed for such macromolecular contacts. The Symplekin HEAT region not only structurally aligns with several established scaffolding proteins, but also has been reported to contact proteins essential for regulating 3′-end processing. Together, these data support the conclusion that the Symplekin HEAT domain serves as a scaffold for protein–protein interactions essential to the mRNA maturation process.

Keywords: crystal structure; molecular dynamics; protein scaffold; HEAT repeat; mRNA processing

Edited by J. Doudna

Introduction

Maturation of most eukaryotic pre-mRNAs requires cleavage and polyadenylation of the 3′-ends of primary transcripts. The 3′-end poly(A) tail ensures proper translation by delivering ribosomes to the mRNA;[1] in amphibian oocytes, it was shown that translation was eliminated when the poly(A) tail addition was blocked by chemical modification.[2] The poly(A) tail is also essential for protecting the message from exonucleases and for transporting the message from the nucleus to the cytoplasm.[3] The length of the poly(A) tail affects the stability of the message, and compromised stability has been shown to lead to inflammation, cancer, early developmental maladies and coronary ailments.[4] Thus, proper poly(A) tail addition to messenger RNA is required for proper cellular function.

For polyadenylation to occur, the cleavage stimulation factor (CstF) and the cleavage and polyadenylation specificity factor (CPSF) must work in concert to recognize and orient the cleavage site for the addition of the poly(A) tail.[5] The ~1160-residue Symplekin protein is proposed to be the scaffolding factor on which this large protein complex is assembled.[3] Symplekin binds two members of the CstF macromolecular complex, CstF64 and CstF77, in

*Corresponding author. E-mail address: redinbo@unc.edu.

Abbreviations used: CstF, cleavage stimulation factor; CPSF, cleavage and polyadenylation specificity factor; HSF, heat shock factor; SAD, single-wavelength anomalous diffraction; MD, molecular dynamics; ARM, armadillo; NLS, nuclear localization signal; MBP, maltose-binding protein; APF, atomic position fluctuation; TEV, tobacco etch virus.

a mutually exclusive manner.[6] Symplekin was identified as a stoichiometric component of the polyadenylation complex recently isolated from mammalian cells.[7] Symplekin, CPSF73 and CPSF100 are part of a stable complex in *Drosophila melanogaster* as shown via coimmunoprecipitation and codepletion studies.[8]

Metazoan replication-dependent histone mRNAs are unique in that their 3′-ends are cleaved, but not polyadenylated. Interestingly, fractionation of HeLa cell nuclear extracts also identified Symplekin as a component of the histone pre-mRNA processing machinery.[9] Additionally, an extensive RNA interference screen found Symplekin to be necessary for histone pre-mRNA processing in *D. melanogaster*; when Symplekin was RNA interference depleted, a histone pre-mRNA reporter[10] and endogenous histone mRNA[8] were misprocessed. These data lead to the hypothesis that Symplekin is essential for proper 3′-end formation of canonical and histone mRNA by providing a scaffold on which protein–protein interactions can occur.[6,9]

Symplekin may also serve as a bridging factor between the polyadenylation machinery and transcription regulators. Most recently, the N-terminal region of yeast Symplekin (Pta1) was found to interact with Ssu72, an RNA polymerase II C-terminal domain serine 5-phosphatase.[11] The 124 N-terminal residues of mouse Symplekin interact with heat shock factor 1 (HSF1). HSF1, Symplekin and other polyadenylation factors coimmunoprecipitate with HSF1 after heat shock, leading to the suggestion that HSF1 stimulates both transcription and processing.[12] Overexpression of a non-DNA-binding mutant of HSF1, which can sequester Symplekin, decreased Hsp70 mRNA polyadenylation in stressed cells.[12] Thus, the N-terminal region of Symplekin may be involved in protein–protein interactions that help couple transcription and processing.

With the use of *in silico* methods,[13–19] several potential HEAT repeats were identified in the N-terminus of *D. melanogaster* Symplekin. Protein domains formed by HEAT repeats are established protein–protein interaction scaffolds.[20–27] HEAT repeats are composed of 37–47 residues that fold into two antiparallel helices connected by short (1–10 amino acids) linkers. Each set of helices can repeat 3 to 36 times, creating a HEAT domain.[16] To characterize the N-terminal region of the Symplekins, the three-dimensional structure of *D. melanogaster* Symplekin residues 19–271 was determined by single-wavelength anomalous dispersion (SAD) phasing and refined to 2.4 Å resolution. Additionally, molecular dynamics (MD) simulations were employed to examine motion within this molecular scaffold. Together, these results provide the first detailed structural information on Symplekin and indicate that the Symplekin HEAT domain may serve as a scaffold for protein–protein interactions essential to the mRNA maturation process.

Table 1. Data collection, phasing and refinement statistics

Data collection	
X-ray source	APS SER-CAT BM-22
Space group	$P4_12_12$
Unit cell parameters	68.7, 68.7, 138.5; 90,
a, *b*, *c* (Å); α, β, γ (°)	90, 90
Wavelength (Å)	
Selenomethionine	0.9719
Native	0.97958
Resolution (Å) (highest shell)	
Selenomethionine	50.0–2.9 (3.0–2.9)
Native	50.0–2.4 (2.49–2.40)
R_{sym}	
Selenomethionine	9.4 (34.4)
Native	8.0 (41.9)
I/σ	
Selenomethionine	22.4 (1.0)
Native	24.8 (1.9)
Completeness (%)	
Selenomethionine	78.1 (6.7)
Native	96.1 (79.6)
Redundancy	
Selenomethionine	10.4 (1.6)
Native	6.4 (2.8)
Phasing	
Mean Figure of Merit	
Centric	0.71
Acentric	0.68
All	0.69
Refinement	
Resolution (Å)	50.0–2.4
No. of reflections	12,465
R_{work}	0.2068
R_{free}	0.2653
Molecules per asymmetric unit (AU)	1
No. of amino acids per AU	248
No. of water molecules per AU	142
Average *B*-factors	46.37
RMS deviations	
Bond lengths (Å)	0.0059
Bond angles (°)	1.20
Ramachandran (%)	
Favored	96.76
Outliers	0.40

$R_{sym} = \sum |I - I_{mean}| / \sum I$ where I is the observed intensity and I_{mean} is the average intensity of several symmetry-related observations.
$R_{work} = \sum |F_o - F_c| / \sum F_o$ where F_o and F_c are the observed and calculated structure factors, respectively.
R_{free} is calculated as above for 5% of data not used in any step of refinement.

Results

Structure of the Symplekin HEAT domain

Examination of the 1165-residue *D. melanogaster* Symplekin sequence using secondary-structure prediction algorithms indicated that a series of HEAT repeats is present in the first 300 amino acids of the protein and that this domain was expected to be conserved in Symplekin orthologues.[13–16,19,28] The predicted *D. melanogaster* Symplekin HEAT domain (residues 19–271) was cloned and expressed in *Escherichia coli*, purified to homogeneity and crystallized using hanging-drop vapor diffusion. The structure of the selenomethionine-substituted Symplekin HEAT domain was determined to 2.9 Å resolution using SAD phasing methods , and the structure of the native Symplekin HEAT domain