Published in final edited form as:
J Mol Biol. 2009 September 11; 392(1): 115–128. doi:10.1016/j.jmb.2009.06.062.

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

Abstract

The majority of eukaryotic pre-mRNAs are processed by 3′-end cleavage and polyadenylation, although in metazoa the replication-dependant 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 ~1,160-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 using SAD phasing methods. The structure exhibits 5 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. Taken 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

Address correspondence to: Matthew R. Redinbo, Ph.D., Department of Chemistry, CB #3290, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3290 (919) 843-8910; Fax (919) 962-2388; E-mail: redinbo@unc.edu.

ACCESSION CODE

Structural coordinates have been deposited with the RCSB PDB as 3GS3.

INTRODUCTION

Maturation of most eukaryotic pre-mRNAs requires cleavage and polyadenylation of the 3′-ends of primary transcripts. The 3′-end polyA tail ensures proper translation by delivering ribosomes to the mRNA[1]; in amphibian oocytes, it was shown that translation was eliminated when the polyA tail addition was blocked by chemical modification[2]. The polyA 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 polyA 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 polyA 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 polyA tail[5]. The ~1,160 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 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 *D. melanogaster* as shown via co-immunoprecipitation and co-depletion 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 (RNAi) screen found Symplekin to be necessary for histone pre-mRNA processing in *D. melanogaster*; when Symplekin was RNAi-depleted, a histone pre-mRNA reporter[10] and endogenous histone mRNA[8] was 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 (CTD) 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]. Over-expression 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.

Utilizing *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 anti-parallel 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 using SAD phasing and refined to 2.4 Å resolution. Additionally, molecular dynamics simulations were employed to examine motion within this molecular scaffold. Taken 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.

RESULTS

Structure of the Symplekin HEAT Domain

Examination of the 1,165 residue *D. melanogaster* Symplekin sequence using secondary structure prediction algorithms indicated that a series of HEAT repeats are 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 *E. coli*, purified to homogeneity and crystallized using hanging-drop vapor diffusion. The structure of the selenomethionine-substituted Symplekin HEAT domain was determined using SAD phasing methods to 2.9 Å resolution, and the structure of the native Symplekin HEAT domain was then refined to 2.4 Å resolution (Table 1). Figure 1a illustrates a portion of the Symplekin HEAT domain final model in the original 2.9 Å resolution experimental density from SAD phasing. Residues 19–271 of *D. melanogaster* Symplekin contain five HEAT repeats that fold into a single domain with a crescent shape (Figure 1b). The ten HEAT helices (residues 22–256) are lettered conventionally for HEAT repeat domains (A for the convex and B for the concave surfaces). Repeats 1–5 contain 37, 37, 47, 46, and 42 amino acids, respectively, values similar to those observed for established HEAT repeats[29].

An extended 31-residue loop (amino acids 187–217 and denoted loop 8) connects helices 4B and 5A in the Symplekin HEAT domain structure. Five polar interactions are formed between this loop and helices 4B and 5A, as well as two internal hydrogen bonds that occur between residues within the loop (Figure 1c). Specifically, within the loop, a 2.8 Å hydrogen bond is formed between the main-chain nitrogen of D192 and the side-chain oxygen of S195, and a 2.9 Å hydrogen bond is observed between the main-chain nitrogen of S203 and a D206 side-chain oxygen. Between the loop and the canonical HEAT domain scaffold, hydrogen bonds are observed between R258 of loop 10, M257 of α5B, K132 of α3B, and residues S195, G200, D201, and S203 of loop 8. Figure 2 illustrates the electrostatic potential of the concave surface of the molecule, indicating the presence of a positively charged patch as well as the predominantly negatively charged loop 8. The average thermal displacement parameter (B-factor) for loop 8 is 69 $Å^2$, while the overall average B-factor for the structure is 52 $Å^2$. One crystal contact involving loop 8 exists in the refined crystal structure, between D209 in loop 8 and E69 of α2A in a symmetry-related monomer.

Conservation in Symplekin Orthologues

In addition to the reported similarity between amino acids 300–800 of human and yeast Symplekin[6], the HEAT repeats within the N-terminal regions of Symplekin orthologues, including the residues on the concave surface and loop 8, are reasonably well conserved. Figure 3 presents a sequence alignment of the N-terminal ~300 residues of Symplekins from eight representative species: *Drosophila melanogaster*, *Homo sapiens*, *Xenopus laevis*, *Strongylocentrotus purpuratus*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Schizosaccharomyces pombe*, and *Saccharomyces cerevisiae*.[30] While only six amino acid positions (119, 152, 179, 180, 251 and 258) are 100% identical within the domain, 38% are highly similar (defined as 6 or more species containing a similar amino acid type). Of these similar residues, 75% are nonpolar and map to positions in the hydrophobic core of the *D. melanogaster* HEAT domain structure.

When comparing only the three sequences most closely related to *D. melanogaster* (*H. sapiens*, *X. laevis* and *S. purpuratus*), it was found that 28 residues that are completely conserved in the hydrophobic core (Figure 4a; see also Figure 3). In addition, in considering the concave, convex and loop regions of Symplekin, it is evident that the majority of identical residues fall on the concave surface and within the loops (Figures 4b, 4c). The sixteen conserved residues found

on the concave surface account for >20% of the total conserved residues in this HEAT domain (Figure 4b, yellow). Loop regions projecting from the concave surface account for ten conserved residues, five of which are in loop 8 (Figure 4b, cyan). In contrast to the concave side, the convex surface contains only three conserved residues (Figure 4c, green). These data indicate that the HEAT domain is likely conserved in the N-terminal regions of the Symplekins of known sequence, and that the hydrophobic core, concave surface and loop 8 are the regions most highly conserved.

Although sequence variation exists at many positions in the more distant species (sequences in grey, Figure 3), examination of secondary structure predictions indicates that the helical HEAT-like fold is preserved in each putative Symplekin orthologue[31]. All seven sequences have unstructured regions aligning with *D. melanogaster* loops, including the extended loop 8 (underlined in Figure 3). While *S. cerevisiae* secondary structure predictions in the regions of α2B and α4A include sequence inserts, homology modeling with PHYRE[31] and Insight II (Accelrys Software, Inc., San Diego, CA, USA) supports the conclusion that this protein adopts a HEAT-like repeat structure. Taken together, these data indicate that the orthologue sequences shown in Figure 3 are likely to resemble the α-helical *D. melanogaster* Symplekin HEAT domain structure.

Symplekin HEAT Repeats are Classified with Scaffolding Proteins

The closely related HEAT and armadillo structural domains have been sub-classified based on specific amino acid sequences that coincide with functional categorization[20]. To further characterize the Symplekin's N-terminal domain, each of the repeats were structurally aligned and the sequences were compared to the sequence classifications for three types of HEAT sequences (ADB, AAA and IMB), as developed by Andrade *et al.*[20]. The AAA, ADB, and IMB HEAT classes all exhibit a similar pattern of hydrophobic residues and contain conserved residues D19 and R/K 25 near the intrahelical loop, while the sequence logo of the ADB class also contains D/N21 and V/I24[20]. *D. melanogaster* Symplekin contains the ADB pattern: HEAT repeat 2 contains D77, N79, V92, and K83, HEAT repeat 3 includes D114, N115, I120, and K121, while HEAT repeat 4 contains 167D, 170N, 173I and R174. Terminal HEAT repeats are more difficult to classify because they have a different set of packing constraints[20]. The highly conserved P11 of the AAA and IMB classes is lacking in the ADB class, and is also lacking in the HEAT repeats 2, 3 and 4 of Symplekin. Taken together, the residues in the three central Symplekin HEAT repeats indicate that Symplekin may belong to a small ADB subclass of HEAT repeats, a family containing mainly α, β-adaptin and β-coat proteins that function as scaffolds for protein binding and transport. This sub-classification supports the hypothesis that the Symplekin HEAT domain has a structure appropriate for protein-protein interactions.

Symplekin HEAT Structurally Aligns with Protein-Binding Scaffolds

The structure of the *D. melanogaster* Symplekin HEAT domain was examined using Dali to identify proteins of similar structure[32]. While nearly 200 protein structures exhibited homology with the Symplekin HEAT domain, the closest structural neighbors were serine/threonine-protein phosphatase 2A PR65/A subunit (PDB 1b3u), Cullin-associated protein Cand1 (PDB 1u6gc), and karyopherin-α (PDB 1ee4), all of which have HEAT or armadillo (ARM) repeats. Experimental evidence indicates that their HEAT/ARM repeats are involved in protein-protein interactions and the majority of these domains utilize their concave face as a protein binding or scaffolding surface[22,24,27,33–36]. Recall that amino acid conservation supported the functional importance of the concave surface and loop 8 of the Symplekin HEAT domain (see two previous sections). Symplekin superimposes on the structure of Cand1 (TIP120) of the Cand1-Cul1 complex with only 10% sequence identity but with 3.8 Å RMSD over 203 aligned residues, and a Z-score of 14.7 (Figure 5a). The concave surface of Cand1 is employed in binding Cul1 to inhibit Cul1 from forming the E3 ubiquitin ligase complex[22]. Symplekin

structurally superimposes on yeast karyopherin-α with 11% identity over 196 Cα positions, a 5.0 Å RMSD, and a Z-score of 14.2 (Figure 5b)[32]. The concave surface angles of each protein were calculated by measuring the angle between three concave surface Cα residues on helices 1B, 3B and 5B at three positions on these helices: near the N-terminus, the center, and near the C-terminus. The concave surface angle for the helical N-termini of Symplekin, Cand1 and karyopherin-α are 144°, 100°, and 153°, respectively; for the helical centers are 141°, 124°, and 157°, respectively; and for the helical C-termini are 107°, 137°, and 150°, respectively. The twist of each HEAT domain was determined by comparing the angle between the helical axes of helices 1B and 5B, and found to be 5°, 10°, and 77° for Symplekin, Cand1 and karyopherin-α, respectively. Thus, while the overall concave surface angles of each HEAT domain are similar, Symplekin and Cand1 exhibit significantly less domain twist than does karyopherin-α.

The core of *S. cerevisiae* karyopherin-α is a canonical ARM repeat with acidic concave surface regions equipped to bind basic nuclear localization signals (NLS)[36]. It has been reported that *D. melanogaster* karyopherin-α3 binds to the positively charged NLS of HSF1[37], and residues 1–124 of mouse Symplekin interact with HSF[12]. *D. melanogaster* and *S. cerevisiae* karyopherin-α sequences share 50% identity and maintain a similar electrostatic surface. There are no extended loops in either karyopherin-α sequence. However, with respect to loop 8, it is clear from the structural superposition of karyopherin-α and the Symplekin HEAT domain that the position of loop 8 clashes with the NLS sequence bound to the surface of karyopherin-α (Figure 5b). Loop 8 of Symplekin is negatively charged and could provide an alternative binding region for the positively charged NLS (Figures 2, 5b); however, the exact region of HSF1 that binds to Symplekin has yet to be determined. Taken together, the observations that karyopherin-α3 and Symplekin contain similar structural motifs, have similar electrostatic surfaces, and both bind to HSF1 support the conclusion that Symplekin has characteristics of a protein-binding scaffold.

Loop 8 Impacts Symplekin HEAT Domain Motion

Molecular dynamics (MD) simulations have been used to investigate the manner in which HEAT and ARM domains change conformational states upon ligand binding, and to design ideal ARM domains for general peptide binding[38–40]. Our attempts at biochemically characterizing the interactions between the Symplekin HEAT domain with *D. melanogaster* CstF64 and Ssu72 through amylose-affinity pull down assays were unsuccessful due to non-specific interaction with the MBP tag. However, these interactions have been shown indirectly in Symplekin orthologues. Instead, we employed MD to examine how loop 8 impacts the overall and correlated motions within the Symplekin HEAT domain structure. Three models of the Symplekin HEAT domain were subjected to 10 ns MD simulations: Wild-Type containing a complete loop 8, a model in which the ten polar residues in loop 8 were all replaced with serine (Poly-Ser Loop 8; serine was chosen to maintain polarity by using small polar side chains in this surface-exposed loop), and a model in which loop 8 is replaced with a short canonical HEAT turn (Short Loop 8) (Figure 6). The Short Loop 8 mutant was designed with the intention of mimicking the minimal loops commonly seen between HEAT repeats. Comparing the Short Loop 8 with Wild-Type Symplekin was expected to show the role loop 8 plays in the motion of the Symplekin HEAT domain. The Poly-Ser Loop 8 model was expected to show whether specific residues on the loop were important for Symplekin HEAT domain motion.

Simulations of each Symplekin model were performed in triplicate using different random number generator seeds. Data used for analysis of each individual simulation was collected from 10 consecutive nanoseconds of the same conformational ensemble (designated by a consistent root mean square deviation from the starting crystal structure) (Figure 7a). The

models were analyzed with respect to both the overall degree of motion seen in Cα atoms (observed as the atomic position fluctuations (APF) of each Cα) as well as the behavior of each Cα with respect to all other Cα atoms. Wild-Type loop 8 and Poly-Ser loop 8 simulations exhibit nearly identical overall motion in loop 8 Cα atoms as well as throughout the entire protein (Table 2). The similarity of the mean APF between the Wild-Type loop 8 and the Poly-Ser loop 8 indicates that the specific amino acids in loop 8 do not contribute significantly to the overall motion in the HEAT domain. The Short Loop 8 simulation's overall degree of motion was also found to be similar to both the Wild-Type loop 8 and Poly-Ser loop 8 simulation, indicating that the presence of the extended loop 8 does not significantly influence the overall motion of the HEAT domain.

Correlation-anticorrelation plots, which provide information on the relative motion of each residue pair during an MD trajectory, were then generated for these HEAT domain simulations. In Figures 7b–d**,** red indicates correlated motion between two Cα positions (*e.g.*, motion in the same direction), blue indicates anti-correlated motion (*e.g.,* in the opposite direction), and yellow indicates no correlation in motion (two residues that move randomly with respect to one another). Both the Wild-Type and Poly-Ser Symplekin HEAT domain simulations exhibit similar patterns and levels of correlated and anticorrelated motion (Figures 7b, 7d), indicating that the dynamics of the HEAT domain is maintained regardless of the specific residues present in loop 8. In contrast, however, the Short loop 8 simulation exhibits noticeably higher levels of correlated and, particularly, anticorrelated motions (Figure 7c), indicating that removal of the loop increases the degree of specific residue-to-residue motions within the HEAT domain. Taken together, these results indicate that the presence of loop 8, but not specific polar residues on the loop, reduces specific pairwise motions in the Symplekin HEAT domain. Thus, maintaining an extended loop in this location in Symplekin (*e.g.*, see Figure 3) may disrupt specific domain movements to provide the neutral scaffold for protein-protein interactions.

DISCUSSION

The 1165-residue Symplekin protein is a component of the 3′-end processing machinery critical to both canonical and histone messenger RNA[3,6,9]. While structural information is available for many of the other 3′-end processing factors[41–49], no structures have been reported for any region of Symplekin from any species to date. Here, we show that residues 19–271 of *D. melanogaster* Symplekin fold into a HEAT domain structure with an extended loop 8 that appears to be conserved in the Symplekins of known sequence. Examination of the electrostatic potential of the Symplekin HEAT domain reveals that the concave surface is positively charged, while the ridge formed by the even-numbered loops exhibits a slight overall negative charge (Figure 2). Sub-classification of Symplekin's HEAT repeats and structural alignments indicate that these regions of Symplekin may act as a scaffold for protein-protein interactions. Indeed, HEAT domains are well established platforms for macromolecular complex formation (*e.g.,*Figure 5). For example, crystal structures and molecular dynamics studies of importin-β reveal four regions for peptide binding within 5 HEAT repeats[50].

Takagaki *et al*. has reported that the central region (residues 300–740) of human Symplekin is 31% similar to *S. cerevisiae* Symplekin orthologue, Pta1[6]. Using the crystal structure reported here as a guide, we further examined Symplekin orthologue sequences and have found that the N-terminal region of Pta1 exhibits some homology to the equivalent region of *D. melanogaster* Symplekin, which is clearly orthologous to human Symplekin. All of the orthologues contain similar hydrophobic/hydrophilic residue distributions and have greater than 60% α-helical content in their N-terminal regions (Figure 3). Specifically, Pta1 maintains hydrophobic residues in 79 out of the 125 hydrophobic positions present in the *D. melanogaster* N-terminal HEAT domain and 33 residues are identical between these two species. Loop 8 lacks secondary structure in all species investigated, and three residues are

identical and nine residues are similar between Pta1 and *D. melanogaster* Symplekin within this 31-residue region. These data support the conclusion that the N-terminus of *S. cerevisiae* Pta1 likely encodes an α-helical HEAT-like domain similar to the *D. melanogaster* HEAT domain structure reported here.

Several published reports map specific protein docking sites within the Symplekin HEAT region. The HEAT domain of the yeast Symplekin homologue Pta1 has been shown to bind to both Ssu72 and Glc7[11,51] and a portion of the mouse Symplekin HEAT domain interacts with HSF1[12]. Ssu72 and Glc7 have been implicated in the regulation of 3′-end processing. Depletion of the Glc7 phosphatase causes an accumulation of phosphorylated Pta1 and a subsequent reduction in 3′-end polyadenylation; this effect can be rescued by the addition of either Glc7 or unphosphorylated Pta1 back into the processing reaction[51]. The binding of yeast Symplekin homologue Pta1 to Ssu72, an RNA polymerase II C-terminal domain phosphatase, may position the 3′-end processing machinery in proximity to primary transcripts to promote facile processing[11]. Similarly, Symplekin may link 3′-end processing to transcriptional control via contacts with transcription factors like HSF1. The binding of the HEAT domain of mouse symplekin to HSF1 promotes polyadenylation of Hsp70 mRNA in heat stressed cells[12]. Taken together, these data indicate that the Symplekin HEAT region provides a platform for enzymes and other proteins critical to modulating 3′-end processing.

GST pull down studies and yeast two-hybrid assays provide information on the specific Symplekin regions involved in these protein-protein interactions (Figure 8). Binding to Glc7 was maintained using Pta1 Δ1–100, while removal of Symplekin residues 1–200 abolished Glc7 binding[51]. Indirectly, this indicates Symplekin HEAT repeats 3, 4 and loop 8 (Pta1 residues 100–200) are used in binding to Glc7[51]. Ssu72 requires Symplekin HEAT repeat 2 for optimal binding (Pta1 residues 51–76)[11], and HSF1 binds to residues HEAT repeats 1–3 (mouse Symplekin 1–124)[12,51]. The exact regions of Symplekin required for interacting with the core 3′-end processing machinery, CstF and CPSF, have not been determined; however, it has been shown that some processing in yeast can occur with a Δ1–300 Pta1 construct[11]. Therefore, we propose a model where several regulatory proteins bind in a mutually exclusive manner to distinct sites on the Symplekin HEAT domain, whereas the C-terminal region of the protein associates with central members of the 3′-end processing machinery (Figure 8).

Molecular dynamics simulations conducted on the *D. melanogaster* Symplekin domain structure provides preliminary insight into the motions in this region of the protein. Although the timescale on the trajectories were only 10 nsec and the domain was examined in isolation, it was clear that the loop 8 is involved in disrupting the dynamic relationship between the residues across the entire HEAT domain (Figure 7b–d). These observations suggest that the wild-type Symplekin HEAT domain may be tuned to adopt a more neutral range of motions to prepare it for binding to different protein partners. Changes in flexibility of wild-type proteins relative to specific mutants have been reported in previous molecular dynamics studies[52]. Additionally, our characterization of the Symplekin HEAT domain agrees well with published MD studies of Armadillo and HEAT domain proteins. Examination of Cse1p by MD indicates that a particularly negatively charged loop (insert 19) helps to poise the structure in an open conformation to facilitate binding to RanGTP and Kap60p[39]. Loop 8 in *D. melanogaster* Symplekin also exhibits a slight overall negative charge and may play a similar role in preparing the domain to bind to protein partners Glc7, Ssu72 or HSF1. In simulations of importin-β, the ligand bound states are curved in shape, but upon ligand release the domain opens to produce more elongated states[39,40]. The Symplekin HEAT domain may also employ such “tertiary disorder”[40] in conforming to different protein-binding partners. It is likely that there may be additional partners that interact with the HEAT domain that may be involved in regulating histone pre-mRNA processing.

Combining our structural and molecular dynamics results with biochemical studies, we have classified the Symplekin HEAT domain as a scaffold for the binding of proteins critical to modulating 3′-end mRNA processing. Utilizing sequence conservation data (Figures 3, 4), future biochemical and mutagenesis studies will be conducted with this HEAT domain to identify specific residues vital for binding to Glc7, Ssu72 and HSF1. A preliminary cryo-EM image of the purified 3′-end processing complex including Symplekin, CPSF, CstF and CFI has recently been determined at low resolution[7] and crystal structures exist for several components of the eukaryotic 3′-end processing machinery, including CPSFs 30, 73, 100, CstF64 and 77 and CFI_m-25[41,42,46,47]. Thus, a range of efforts are underway to understand the intricate macromolecular relationships required for the catalytic and regulatory aspects of 3′-end processing machinery. The structure of the Symplekin HEAT domain presented here provides an additional piece of this complex structural puzzle.

MATERIALS AND METHODS

Expression and Purification of Symplekin HEAT Domain

The following software programs were utilized to predict the structural elements within Symplekin: BLAST, Jpred[15], PHYRE[17], pFam[16], InterProScan[19], ScanSite[53], PredictProtein[18], RONN[54] and COILS[55]. The disordered regions include 1–18, 452–544, and 1116–1165. A HEAT-like domain was predicted between residues 19–271. Based on these analyses, residues 19–271 of *D. melanogaster* Symplekin were cloned into the expression vector pMCGS9, which provided N-terminal 6-histidine and maltose-binding protein (MBP) tags followed by a Tobacco Etch Virus (TEV) protease site[56]. *Escherichia coli* BL21 (DE3) gold cells (Stratagene) were transformed with this constructed plasmid and cells were grown at 37 °C in 1.5 L of terrific broth supplemented with 50 mg/L ampicillin until an A_{600}=1.0–1.2. The temperature was dropped to 18 °C and 0.1 mM of IPTG was added to induce protein expression until a final OD A_{600}=4.5. The cells were harvested by centrifugation and resuspended in nickel buffer A (5 mM imidazole, 50 mM potassium phosphate, pH 7.4, 150 mM NaCl, 1 mM DTT, 0.01% sodium azide) and stored at −80°C. Thawed cells were lysed by sonication in the presence of DNase and protease inhibitors, and centrifuged at high speed for 60 minutes to produce a cleared lysate. The histidine-tagged protein was purified from the lysate by nickel affinity chromatography. Nickel buffer B (500 mM imidazole, 50 mM potassium phosphate, pH 7.4, 150 mM NaCl, 1 mM DTT, 0.01% sodium azide) was used to elute the protein from the column with a gradient of 5–100% B. To cleave the 6xHis-MBP fusion protein from the Symplekin 19–271 polypeptide, 2% TEV protease by mass TEV/mass Symplekin was added. Protein was dialyzed into nickel buffer A during TEV cleavage. A second nickel column purified the now un-tagged Symplekin from the 6×His-MBP tag. A polishing step of size exclusion chromatography (Column: Superdex 75, GE Healthcare; sizing buffer: 10 mM HEPES, pH 8.0, 50 mM NaCl, 1 mM DTT and 0.01% sodium azide) produced >95% purity by SDS PAGE. A selenomethionine-substituted form of *D. melanogaster* Symplekin residues 19–271 was produced using B834 cells, a methionine auxotroph cell line. Cells were grown in selenomethionine specific media (Athena) supplemented with 50 mg/L selenomethionine. Expression and purification procedures were identical to those listed above for the native protein.

Crystallization and X-ray Data Collection

Native and selenomethionine-substituted Symplekin proteins were concentrated to 3–6 mg/mL in sizing buffer. Crystallization was performed by hanging drop diffusion at 22 °C with mother liquor consisting of 0.4–0.5 M sodium citrate, 25–28% PEG 3350, 10 mM HEPES, pH 8.0, 0.01% N_3Na and 1 mM DTT. Each crystallization drop contained 1 μL of protein and 1 μL of well solution. Diamond shaped crystals grew within one week, with maximal dimensions of 300 μm × 60 μm × 60 μm. Crystals were cryoprotected in mother liquor plus 35% PEG 3350

and flash-cooled in liquid nitrogen. Diffraction data were collected at 100K using Sector 22-BM (SER-CAT) of the Advanced Photon Source at Argonne National Laboratories. A SAD data set was collected on crystals containing selenomethionine-substituted protein at 0.97190 Å; a native data set was collected using crystals containing wild-type protein at 0.97958 Å. DENZO and SCALEPACK in HKL-2000 were employed for data indexing and scaling[57]. The crystals were of the space group $P4_12_12$ with unit cell dimensions of a, b = 68.7 Å, c = 138.5 Å and α, β, γ = 90° (Table 1).

Phasing, Model Building and Refinement

The SGXPRO software package, an interface for programs including SHELXD and SOLVE/RESOLVE, was employed to identify heavy atom sites and provide initial phases[58]. A Matthews's coefficient value of 2.9 indicated that 1 molecule was expected in the asymmetric unit with 57.6% solvent. Six methionine residues were present in Symplekin 19–271, thus six Se sites were expected. SHELXD and SOLVE identified all six Se atom positions, and initial phases were calculated to 2.9 Å. RESOLVE was used for density modification and to provide an initial model. After these steps, the overall figure of merit was 0.69.

The model was built further by hand using COOT[59]. Initially, all helices were built with alanine residues. Loops were added over several rounds of refinement to connect the helices. Finally, side chains were placed in the model. This 2.9 Å model from SAD was refined using REFMAC5 at this stage to R and R_{free} values of 0.353 and 0.419, respectively. To phase the 2.4 Å native data set, the model refined using the SAD data was used in molecular replacement[60]. Further refinement was conducted by building and validating the model in COOT, and employing both CNS and REFMAC5 to produce R and R_{free} values of 0.2068 and −0.2653, respectively (Table 1). For both the original SAD data and the final native data, 5% of the data were set aside for the free-R and not used at any stage of refinement. The final model, consisting of 248 residues (no density was present for residues 19–21 and 271) and 142 water molecules, was validated with PROCHECK and Molprobity[61]. Figures 1, 2, 4, 5 and 6 were created using PyMOL[62].

Sequence and Structural Alignments

The amino acid sequence of *D. melanogaster* Symplekin was entered into NCBI BLAST to retrieve homologous protein sequences. Sequences (with NCBI Accession numbers) from *Drosophila melanogaster* (**NP_649580.1**), *Homo sapiens* (**NP_004810.2**), *Xenopus laevis* (**NP_001079691.1**), *Strongylocentrotus purpuratus* (**XP_783721.2**), *Caenorhabditis elegans* (**NP_505210.2**), *Arabidopsis thaliana* (**NP_195760.1**), *Schizosaccharomyces pombe* (**NP_594351.2**) and *Saccharomyces cerevisiae* (**AAA34919.1**) were selected to represent a broad spectrum of species containing Symplekin. The sequence alignment, prepared using ClustalX and refined using several rounds of PSI-BLAST, was abbreviated to display only the portion of the sequences that align with the *D. melanogaster* HEAT domain structure (Figure 3). The structural alignment shown in Figure 5 was prepared using Dali[32]. To characterize the HEAT repeats, each Symplekin HEAT repeat was structurally aligned to HEAT repeat 2.

Molecular Dynamics Simulations

COOT[59] was utilized to create the Symplekin modeled Short loop 8 and Poly-Ser loop 8. For the Poly-Ser model, the residues changed to serine were D192, E193, D194, K197, R198, D199, D201, D209, H210, R215. To design a short turn to replace loop 8 in the Short loop model, many other HEAT repeat proteins were examined to identify common linkers and it was determined that six residues are sufficient to bridge a 10.6 Å gap. To keep this linker as authentic as possible, residues on each end of the loop were maintained and connected with a glycine, a residue common in loops of HEAT repeats. Native residues 190–214 were

completely removed. Thus, the modeled Short loop 8 is 187-LQSGRR-216. Residues 187–216 were used to calculate the relative APF values for loop 8 in each simulation.

Molecular dynamics simulations of the Symplekin HEAT domains were performed in triplicate using the AMBER 2003 force field with at 2 fs time step[63]. LEaP was used to generate the topology and parameter files, SANDER performed the 5000 steps of energy minimization, which included constant volume followed by constant temperature equilibration, the PMEMD module was used for the production runs, and PTRAJ was utilized for analysis of the results[63]. TIP3P water molecules were used to generate the solvated structure[64], and electrostatic interactions were calculated using the particle-mesh Ewald algorithm with a cutoff of 10 Å applied to Lennard-Jones interactions[65]. All molecular dynamics simulations were conducted and analyzed as described previously[66].

Acknowledgments

This study was financially supported by NIH R01 grant DK62229 (M.R.R.) and GM58921 (W.F.M.) and the Graduate Assistants in Areas of National Need (GAANN) fellowship (S.A.K.). M.S. was supported by NIH F32 GM080950. Data were collected at Southeast Regional Collaborative Access Team (SER-CAT) beam-line 22 at the Argonne National Laboratory Advanced Photon Source. Use of the Advanced Photon Source was supported by the Office of Basic Energy Sciences of the U.S. Department of Energy Office of Science under contract no. W-31-109-Eng-38. We thank E. Ortlund, R. Duronio, D. Tatomer, L. Charlton, J. Orans, B. Wallace, K. Brennaman and A. Tripathy for helpful discussions.

References

1. Preiss T, Hentze MW. Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast. Nature 1998;392:516–520. [PubMed: 9548259]
2. Sachs AB, Sarnow P, Hentze MW. Starting at the beginning, middle, and end: translation initiation in eukaryotes. Cell 1997;89:831–838. [PubMed: 9200601]
3. Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3′-end processing. Cell Mol Life Sci 2008;65:1099–1122. [PubMed: 18158581]
4. Wilusz CJ, Wormington M, Peltz SW. The cap-to-tail guide to mRNA turnover. Nat Rev Mol Cell Biol 2001;2:237–246. [PubMed: 11283721]
5. Dominski Z, Marzluff WF. Formation of the 3′ end of histone mRNA: getting closer to the end. Gene 2007;396:373–390. [PubMed: 17531405]
6. Takagaki Y, Manley JL. Complex protein interactions within the human polyadenylation machinery identify a novel component. Mol Cell Biol 2000;20:1515–1525. [PubMed: 10669729]
7. Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR III, Frank J, Manley JL. Molecular Architecture of the Human Pre-mRNA 3′ Processing Complex. Mol Cell 2009;33:365–376. [PubMed: 19217410]
8. Sullivan KD, Steiniger M, Marzluff WF. A core complex of CPSF73, CPSF100, and Symplekin may form two different cleavage factors for processing of poly(A) and histone mRNAs. Mol Cell 2009;34:322–332. [PubMed: 19450530]
9. Kolev NG, Steitz JA. Symplekin and multiple other polyadenylation factors participate in 3′-end maturation of histone mRNAs. Genes Dev 2005;19:2583–2592. [PubMed: 16230528]
10. Wagner EJ, Burch BD, Godfrey AC, Salzler HR, Duronio RJ, Marzluff WF. A genome-wide RNA interference screen reveals that variant histones are necessary for replication-dependent histone pre-mRNA processing. Mol Cell 2007;28:692–699. [PubMed: 18042462]
11. Ghazy M, He X, Singh BN, Hampsey M, Moore C. The essential N-terminus of the Pta1 scaffold protein is required for snoRNA transcription termination and Ssu72 function but is dispensable for pre-mRNA 3′-end processing. Mol Cell Biol. 2009
12. Xing H, Mayhew CN, Cullen KE, Park-Sarge OK, Sarge KD. HSF1 modulation of Hsp70 mRNA polyadenylation via interaction with symplekin. J Biol Chem 2004;279:10551–10555. [PubMed: 14707147]

13. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402. [PubMed: 9254694]
14. Bennett-Lovsey RM, Herbert AD, Sternberg MJ, Kelley LA. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. Proteins 2008;70:611–625. [PubMed: 17876813]
15. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ. JPred: a consensus secondary structure prediction server. Bioinformatics 1998;14:892–893. [PubMed: 9927721]
16. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A. Pfam: clans, web tools and services. Nucleic Acids Res 2006;34:D247–251. [PubMed: 16381856]
17. Kelley LA, MacCallum RM, Sternberg MJ. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol 2000;299:499–520. [PubMed: 10860755]
18. Rost B, Yachdav G, Liu J. The PredictProtein server. Nucleic Acids Res 2004;32:W321–326. [PubMed: 15215403]
19. Zdobnov EM, Apweiler R. InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics 2001;17:847–848. [PubMed: 11590104]
20. Andrade MA, Petosa C, O'Donoghue SI, Muller CW, Bork P. Comparison of ARM and HEAT protein repeats. J Mol Biol 2001;309:1–18. [PubMed: 11491282]
21. Cho US, Xu W. Crystal structure of a protein phosphatase 2A heterotrimeric holoenzyme. Nature 2007;445:53–57. [PubMed: 17086192]
22. Goldenberg SJ, Cascio TC, Shumway SD, Garbutt KC, Liu J, Xiong Y, Zheng N. Structure of the Cand1-Cul1-Roc1 complex reveals regulatory mechanisms for the assembly of the multisubunit cullin-dependent ubiquitin ligases. Cell 2004;119:517–528. [PubMed: 15537541]
23. Groves MR, Hanlon N, Turowski P, Hemmings BA, Barford D. The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. Cell 1999;96:99–110. [PubMed: 9989501]
24. Lee SJ, Matsuura Y, Liu SM, Stewart M. Structural basis for nuclear import complex dissociation by RanGTP. Nature 2005;435:693–696. [PubMed: 15864302]
25. Matsuura Y, Stewart M. Nup50/Npap60 function in nuclear protein import complex disassembly and importin recycling. Embo J 2005;24:3681–3689. [PubMed: 16222336]
26. Paffenholz R, Kuhn C, Grund C, Stehr S, Franke WW. The arm-repeat protein NPRAP (neurojungin) is a constituent of the plaques of the outer limiting zone in the retina, defining a novel type of adhering junction. Exp Cell Res 1999;250:452–464. [PubMed: 10413599]
27. Sampietro J, Dahlberg CL, Cho US, Hinds TR, Kimelman D, Xu W. Crystal structure of a beta-catenin/BCL9/Tcf4 complex. Mol Cell 2006;24:293–300. [PubMed: 17052462]
28. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C. InterPro: the integrative protein signature database. Nucleic Acids Res. 2008
29. Andrade MA, Ponting CP, Gibson TJ, Bork P. Homology-based method for identification of protein repeats using statistical significance estimates. J Mol Biol 2000;298:521–537. [PubMed: 10772867]
30. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;25:4876–4882. [PubMed: 9396791]
31. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 2009;4:363–371. [PubMed: 19247286]
32. Holm L, Sander C. Mapping the protein universe. Science 1996;273:595–603. [PubMed: 8662544]
33. Xu Y, Xing Y, Chen Y, Chao Y, Lin Z, Fan E, Yu JW, Strack S, Jeffrey PD, Shi Y. Structure of the protein phosphatase 2A holoenzyme. Cell 2006;127:1239–1251. [PubMed: 17174897]
34. Wang X, McLachlan J, Zamore PD, Hall TM. Modular recognition of RNA by a human pumilio-homology domain. Cell 2002;110:501–512. [PubMed: 12202039]

35. Neuwald AF, Hirano T. HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. Genome Res 2000;10:1445–1452. [PubMed: 11042144]
36. Conti E, Kuriyan J. Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha. Structure 2000;8:329–338. [PubMed: 10745017]
37. Fang X, Chen T, Tran K, Parker CS. Developmental regulation of the heat shock response by nuclear transport factor karyopherin-alpha3. Development 2001;128:3349–3358. [PubMed: 11546751]
38. Parmeggiani F, Pellarin R, Larsen AP, Varadamsetty G, Stumpp MT, Zerbe O, Caflisch A, Pluckthun A. Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. J Mol Biol 2008;376:1282–1304. [PubMed: 18222472]
39. Zachariae U, Grubmuller H. A highly strained nuclear conformation of the exportin Cse1p revealed by molecular dynamics simulations. Structure 2006;14:1469–1478. [PubMed: 16962977]
40. Zachariae U, Grubmuller H. Importin-beta: structural and dynamic determinants of a molecular spring. Structure 2008;16:906–915. [PubMed: 18547523]
41. Coseno M, Martin G, Berger C, Gilmartin G, Keller W, Doublie S. Crystal structure of the 25 kDa subunit of human cleavage factor Im. Nucleic Acids Res 2008;36:3474–3483. [PubMed: 18445629]
42. Qu X, Perez-Canadillas JM, Agrawal S, De Baecke J, Cheng H, Varani G, Moore C. The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3′-end processing. J Biol Chem 2007;282:2101–2115. [PubMed: 17116658]
43. Balbo PB, Meinke G, Bohm A. Kinetic studies of yeast polyA polymerase indicate an induced fit mechanism for nucleotide specificity. Biochemistry 2005;44:7777–7786. [PubMed: 15909992]
44. Deo RC, Bonanno JB, Sonenberg N, Burley SK. Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell 1999;98:835–845. [PubMed: 10499800]
45. Perez-Canadillas JM. Grabbing the message: structural basis of mRNA 3′UTR recognition by Hrp1. Embo J 2006;25:3167–3178. [PubMed: 16794580]
46. Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L. Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. Nature 2006;444:953–956. [PubMed: 17128255]
47. Bai Y, Auperin TC, Chou CY, Chang GG, Manley JL, Tong L. Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors. Mol Cell 2007;25:863–875. [PubMed: 17386263]
48. Meinhart A, Cramer P. Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. Nature 2004;430:223–226. [PubMed: 15241417]
49. Noble CG, Beuth B, Taylor IA. Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor. Nucleic Acids Res 2007;35:87–99. [PubMed: 17151076]
50. Isgro TA, Schulten K. Binding dynamics of isolated nucleoporin repeat regions to importin-beta. Structure 2005;13:1869–1879. [PubMed: 16338415]
51. He X, Moore C. Regulation of yeast mRNA 3′ end processing by phosphorylation. Mol Cell 2005;19:619–629. [PubMed: 16137619]
52. Rizzuti B, Sportelli L, Guzzi R. Evidence of reduced flexibility in disulfide bridge-depleted azurin: a molecular dynamics simulation study. Biophys Chem 2001;94:107–120. [PubMed: 11744195]
53. Obenauer JC, Cantley LC, Yaffe MB. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res 2003;31:3635–3641. [PubMed: 12824383]
54. Yang ZR, Thomson R, McNeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics 2005;21:3369–3376. [PubMed: 15947016]
55. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science 1991;252:1162–1164.
56. Donnelly MI, Zhou M, Millard CS, Clancy S, Stols L, Eschenfeldt WH, Collart FR, Joachimiak A. An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. Protein Expr Purif 2006;47:446–454. [PubMed: 16497515]

57. Otwinowski, ZaMW. Processing of X-ray Diffraction Data Collected in Oscillation Mode. In: Carter, C., Jr; Sweet, R., editors. Methods in Enzymology, Macromolecular Crystallography, Part A. Vol. 276. Academic Press; New York: 1997.
58. Fu ZQ, Rose J, Wang BC. SGXPro: a parallel workflow engine enabling optimization of program performance and automation of structure determination. Acta Crystallogr D Biol Crystallogr 2005;61:951–959. [PubMed: 15983418]
59. Emsley, PaC; Kevin. Coot: Model-Building Tools for Molecular Graphics. Acta Crystallographica Section D - Biological Crystallography 2004;60:2126–2132.
60. Collaborative computational project. The CCP4 Suite: Programs for Protein Crystallography. Acta Cryst 1994;D50:760–763.
61. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins 2003;50:437–450. [PubMed: 12557186]
62. Delano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; Palo Alto, CA, USA: 2002.
63. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. J Comput Chem 2005;26:1668–1688. [PubMed: 16200636]
64. Jorgensen W, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
65. Essman UPL, Berkowitz ML, Darden T, Lee H, Pedersen L. A smooth particl mesh Ewald method. J Chem Phys 1995;103:8577–8593.
66. Teotico DG, Frazier ML, Ding F, Dokholyan NV, Temple BR, Redinbo MR. Active nuclear receptors exhibit highly correlated AF-2 domain motions. PLoS Comput Biol 2008;4:e1000111. [PubMed: 18617990]

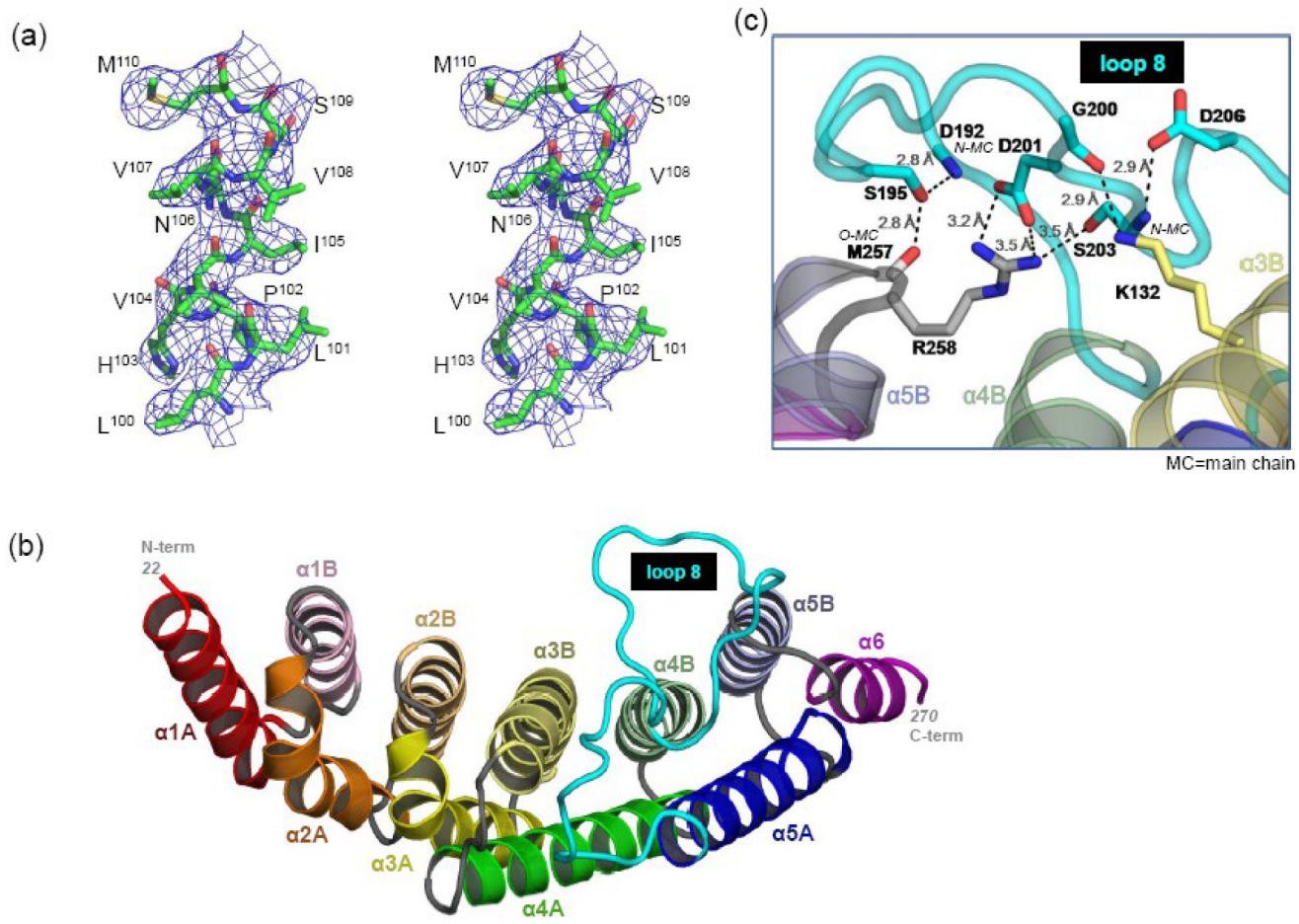

Figure 1.

Symplekin HEAT domain structure. (**a**) A wall-eyed stereo view representation of a portion of the final model in the original experimental electron density from SAD phasing contoured to 1σ. (**b**) The overall structure of the HEAT domain within Symplekin. Helices are lettered and numbered according to classical HEAT naming; the A helices create the convex face, while the B helices create the concave face. The rainbow denotes the N to C progression of residues 22–270. The B helices are in light colors corresponding to their counterpart A helices. For example, 1A is red and 1B is pink. The extended loop 8 is in cyan. (**c**) Polar contacts within the loop 8 region of Symplekin. Arginine 258 and aspartic acid 201 form a salt bridge that anchors loop 8 to helix 5B. Lysine 132 forms a salt bridge with G200 to hold the loop in place with respect to helix 3B. A variety of other polar contacts position the extended loop 8 at the ends of helices 3–5 including S195-D192, M257-S195, R258-S203, S203-D206.

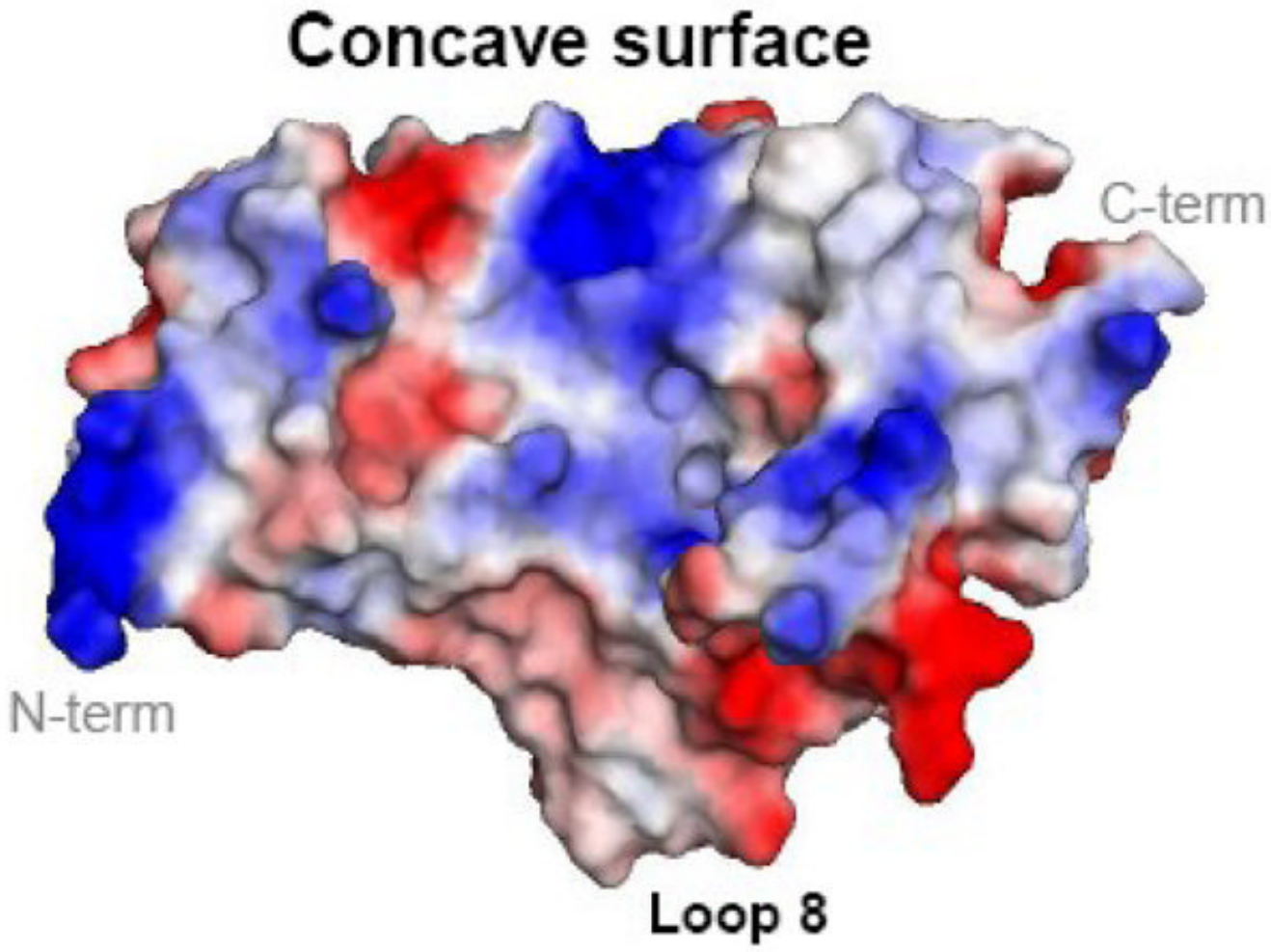

Figure 2.
Electrostatic representation of the concave surface of Symplekin's HEAT domain. Red denotes negatively charged surfaces, blue denotes positively charged surfaces. The molecule is rotated 90° along the horizontal axis of Figure 1b, to orient the concave surface towards the reader. The concave surface is mainly positively charged, while loop 8 is negatively charged.

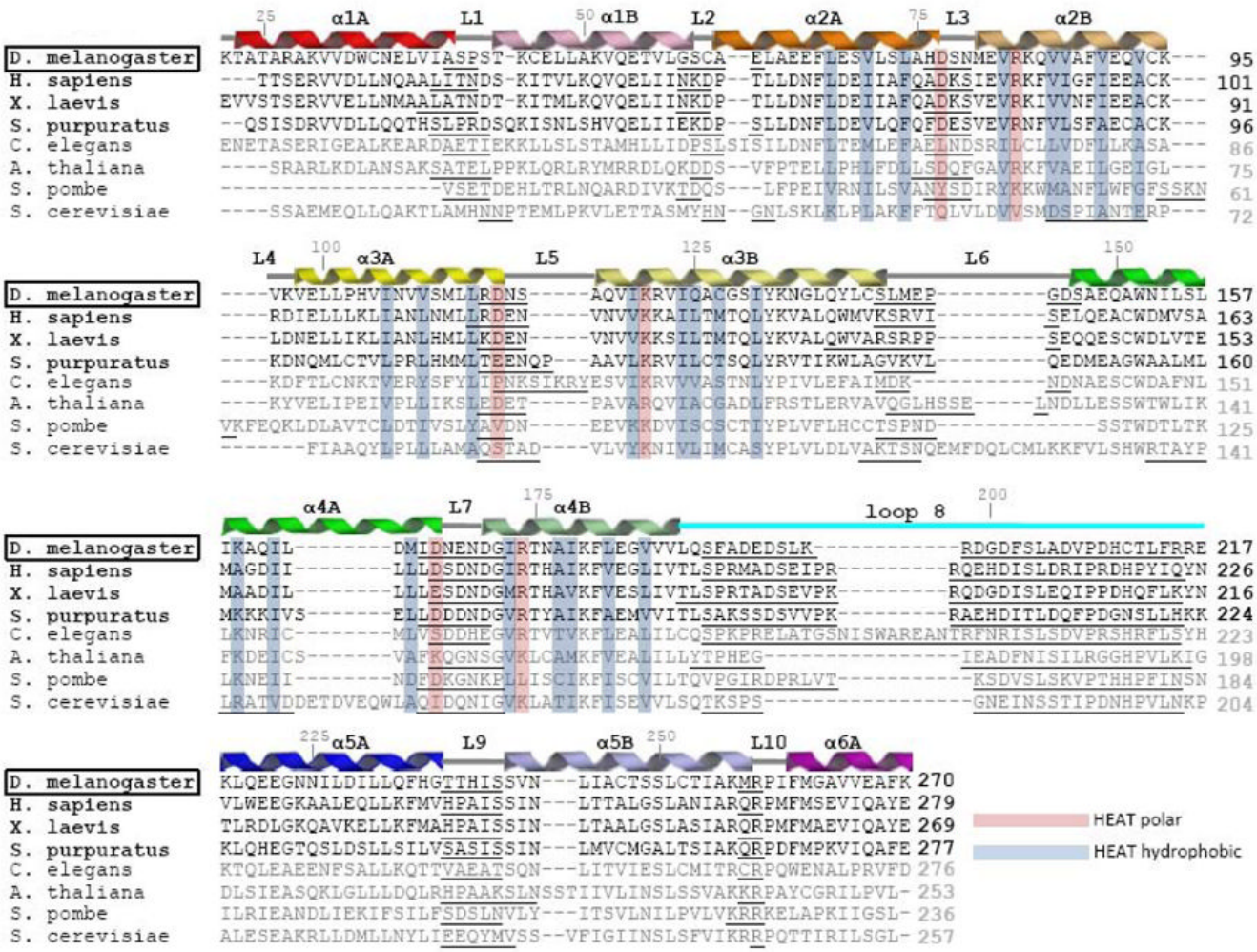

Figure 3.

Sequence alignment of Symplekin orthologues in various species. The secondary structure elements and numbering across the top of the sequences correspond to the *D. melanogaster* structure in Figure 1b. Pink blocks denote the conserved D/E19 and K/R25 required for HEAT repeats, and blue colored blocks represent the HEAT repeat hydrophobic signature. The more distantly related orthologue sequences are shown in grey. Sequences and alignment were made using PSI-Blast and ClustalX. Secondary structure prediction and models of each sequence were predicted using PHYRE. Black underline denotes regions of disorder predicted by PHYRE, non-underline sequences are all predicted to be α-helical.

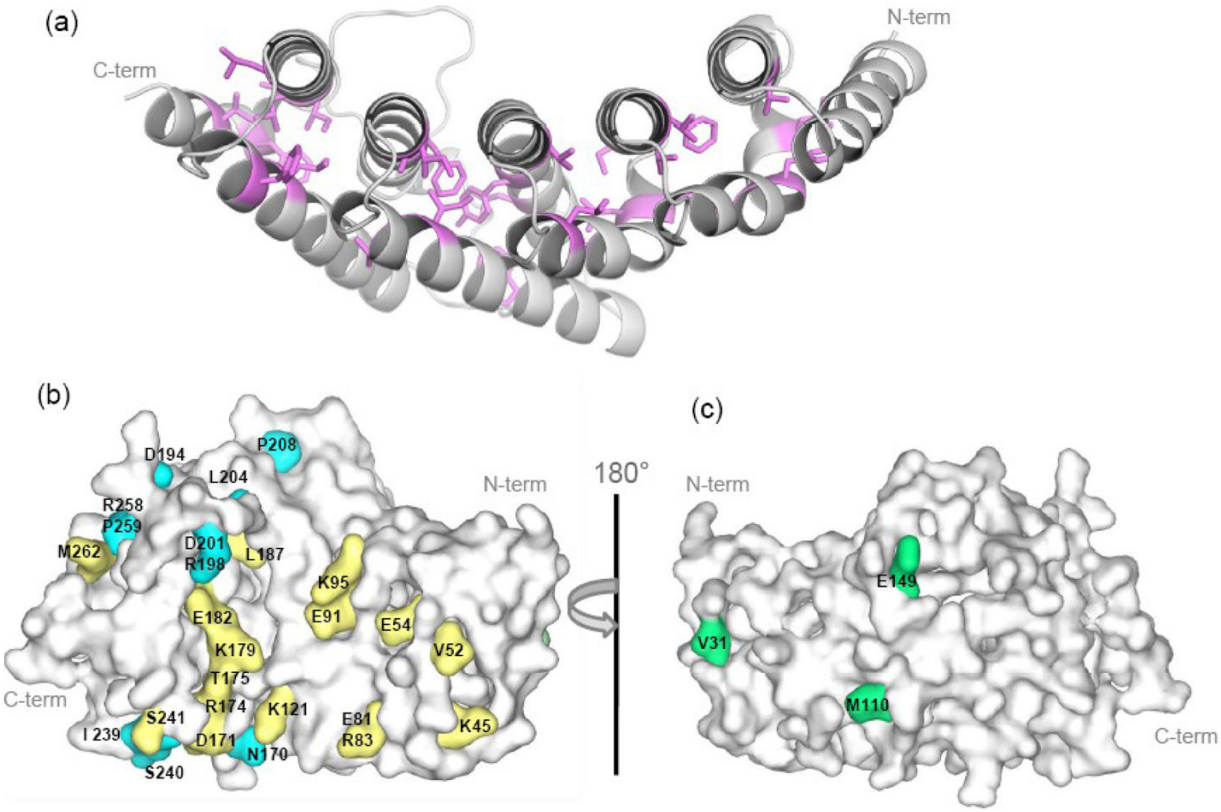

Figure 4.

Conserved residues among four closely related Symplekin orthologues (*H. sapiens, X. laevis, D. melanogaster, S. purpuratus*) mapped onto the HEAT domain structure. (**a**) View of the hydrophobic core where purple represents residues with 100% conservation. (Molecule is rotated 180° on the vertical axis with respect to Figure 1b.) (**b**) View of the concave surface colored as follows: yellow denotes 100% conserved residues that project out of the concave surface, cyan residues are conserved in loop regions, and gray residues are not 100% conserved. (Molecule is rotated 90° on the horizontal axis with respect to Figure 4a.) (**c**) View of the convex surface colored as in A, except green denotes conserved residues that project from the convex surface.

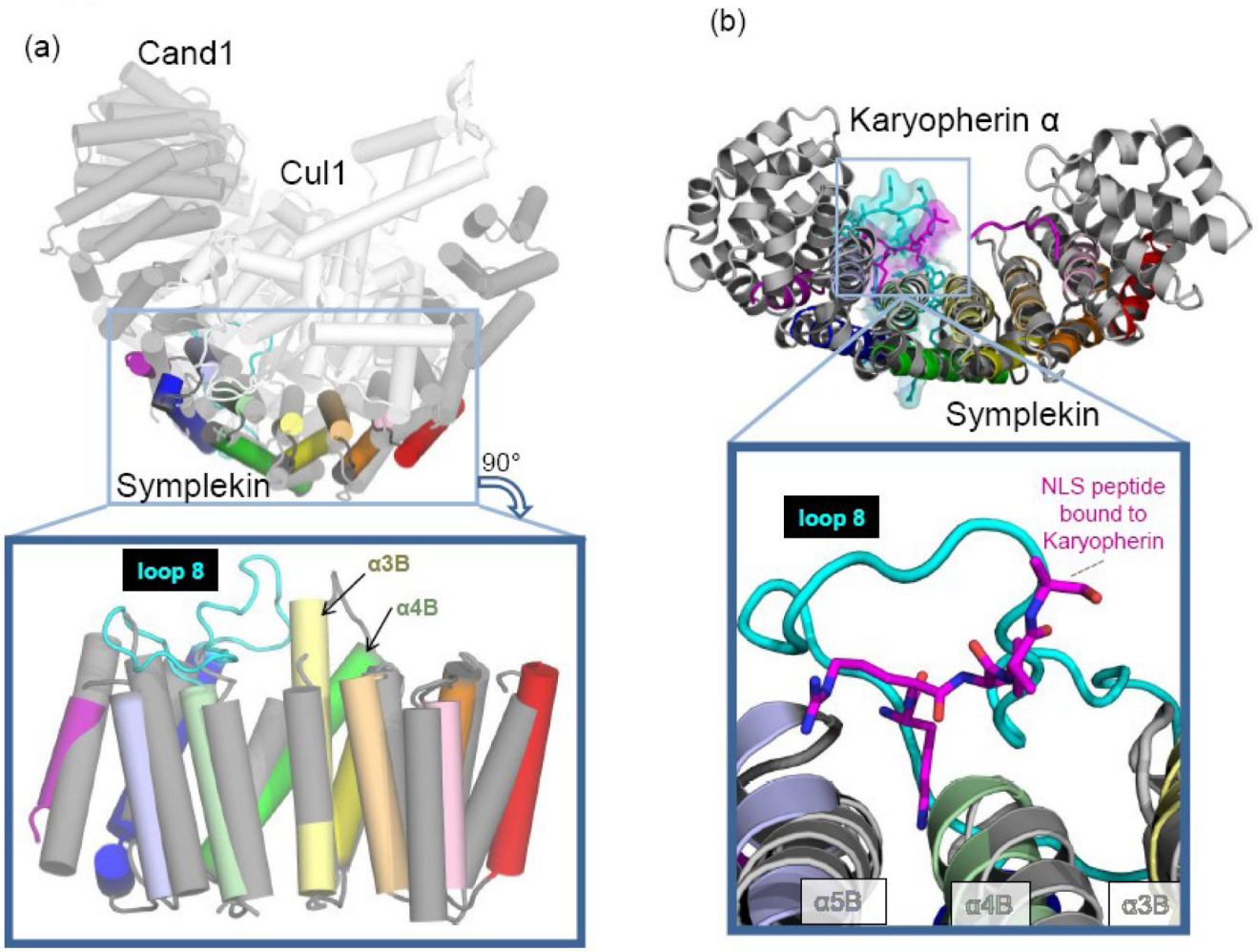

Figure 5.

Symplekin structural alignment with the two most closely related structures. (**a**) Symplekin superimposed with *H. sapiens* Cand1 of the Cand1-Cul1-Roc1 complex (PDB 1u6g). Cand1 structure is in grey, Cul1 in white, and Roc1 is removed for figure clarity. Symplekin helices and surface have coloring from Figure 1b. A closer look at aligned individual helices shows that α3B is extended in comparison to the aligned helix in Cand1. Loop 8 is unique to Symplekin compared to Cand1. (**b**) Symplekin superimposed with *S. cerevisiae* karyopherin-α (PDB 1ee4). Karyopherin-α is grey, the nuclear localization signal (NLS) peptide bound to karyopherin-α is magenta. Symplekin maintains coloring from Figure 1b. Loop 8 docks into α-helices 3B, 4B and 5B, and lies in the same region that the NLS peptide occupies on karyopherin-α.

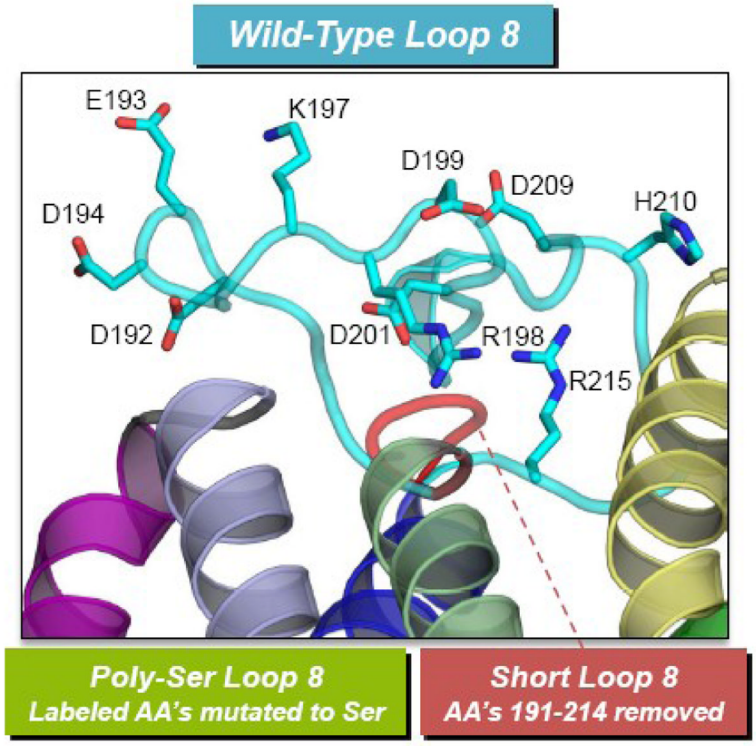

Figure 6.
Symplekin HEAT domain structures used for molecular dynamics simulations. The labeled residues in loop 8 are mutated to serine for the Poly-Ser Loop 8 simulation. To prepare the short loop model, residues 191–214 were removed and wild-type residue 189 was connected to 215 by mutating F190G.

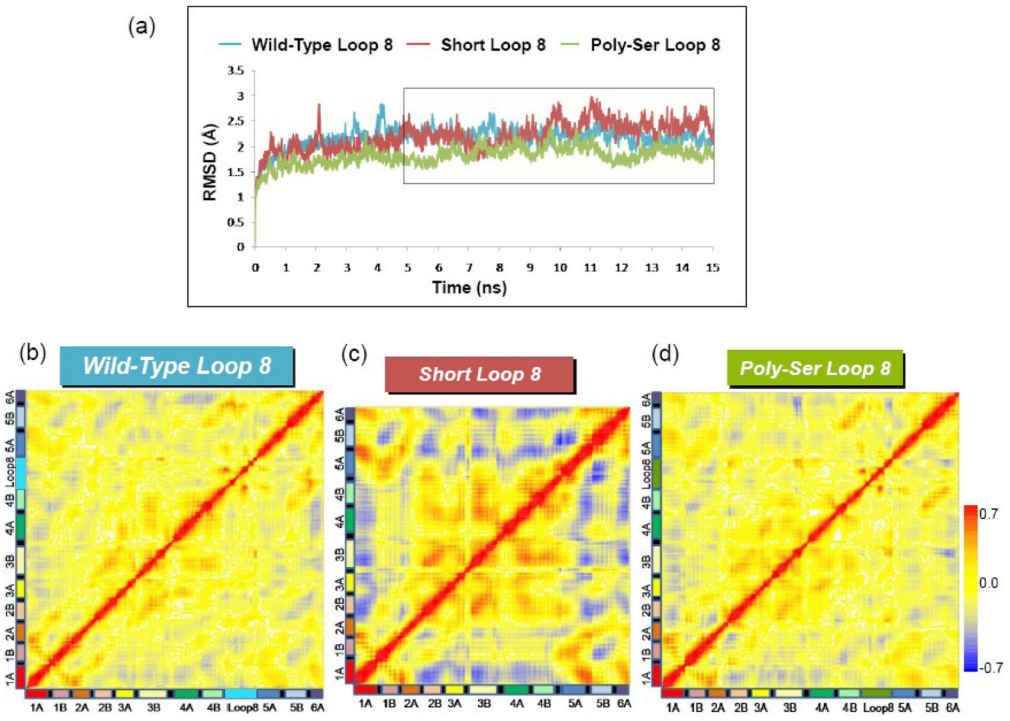

Figure 7.

Truncation of loop 8 increases correlation/anticorrelation within Symplekin's HEAT domain. (a) All atom root mean squared deviation in position over the simulation time scale. The time period between 5–15 ns (boxed region) illustrates a consistent RMSD from the initial model, demonstrating that a stable conformational ensemble was utilized in data analysis. Correlation/anticorrelation plots for Wild-Type (b), Short Loop 8 (c), and Poly-Ser Loop 8 (d) from molecular dynamics simulations. Red represents correlated movement, blue represents anticorrelated movements and colors between represent less correlated movements according to the given color scale. Each axis represents the Cα position for the given residue within the HEAT domain (going from N- to C- terminus from both left to right and from bottom to top). The secondary structural elements are colored consistently with Figure 1b.

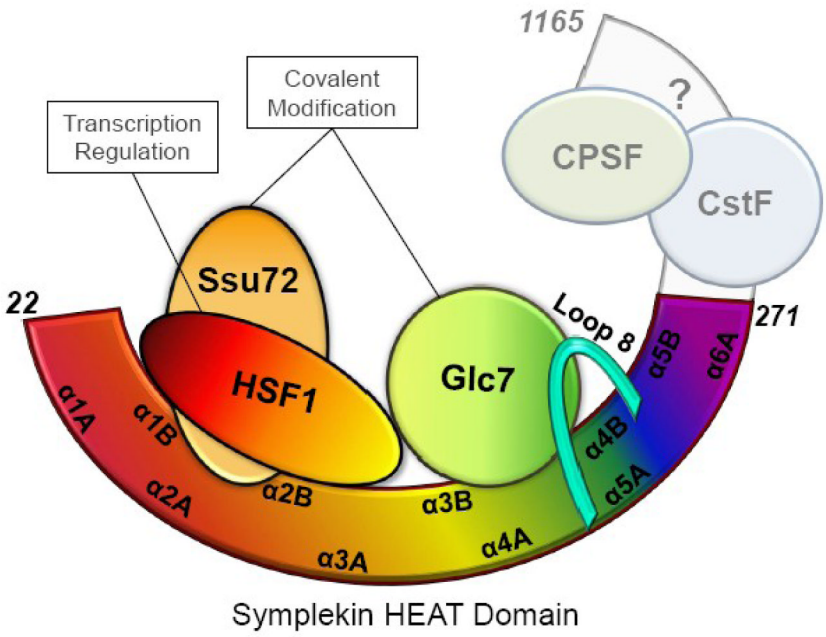

Figure 8.

Symplekin model for protein scaffolding. A diagram illustrating the Symplekin HEAT domain's interaction with known binding partners. Secondary structural elements and residue numbers are labeled according to the structure. HSF1, Ssu72 and Glc7 bind to specific regions of the HEAT domain as described in the text. The C-terminal region of Symplekin has yet to be structurally characterized with respect to binding to the core 3′-end machinery.

Table 1

Data Collection, Phasing and Refinement Statistics

Data collection		
X-ray source	APS SER-CAT BM-22	
Space Group	P41212	
Unit cell a,b,c (Å);α, β, γ (°)	68.7, 68.7, 138.5; 90, 90, 90	
Data set	SeMet	Native
Wavelength (Å)	0.97190	0.97958
Resolution (Å) (highest shell)	50.0–2.9 (3.0–2.9)	50.0–2.4 (2.49–2.40)
R_{sym}	9.4 (34.4)	8.0 (41.9)
I/σ	22.4 (1.0)	24.8 (1.9)
Completeness (%)	78.1 (6.7)	96.1 (79.6)
Redundancy	10.4 (1.6)	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. reflections	12465	
R_{work}	0.2068	
R_{free}	0.2653	
Molecules per asymmetric unit (AU)	1	
No. of amino acids per AU	248	
No. of waters per AU	142	
Average *B*-factors	46.37	
R.M.S. deviations		
Bond lengths (Å)	0.0059	
Bond angles (°)	1.20	
Ramachandran (%)		
Favored	96.76	
Outliers	0.40	

$R_{sym}=\Sigma|I-I_{mean}|/\Sigma I$ where I is the observed intensity and I_{mean} is the average intensity of several symmetry related observations.

$R_{work}=\Sigma|F_o-F_c|/\Sigma F_o$ where F_o and F_c are the observed and calculated structure factors, respectively.

R_{free}=calculated as above for 5% of data not used in any step of refinement.

Table 2

Atomic Position Fluctuations ($Å^2$) for All Cα or Loop 8 Cα Atoms

Model	***Wild-Type Loop 8***		***Poly-Ser Loop 8***		***Short Loop 8***
Residues	**All**	**Loop 8**	**All**	**Loop 8**	**All**
Mean	1.0 ± 0.024	1.2 ± 0.035	1.1 ± 0.087	1.4 ± 0.32	1.2 ± 0.036
Max	3.5 ± 0.57	1.90 ± 0.32	3.8 ± 1.1	1.6 ± 0.095	3.8 ± 0.80
Min	0.48 ± 0.014	0.52 ± 0.036	0.52 ± 0.020	0.54 ± 0.046	0.59 ± 0.031