Structure *14*, 703–712, April 2006 ©2006 Elsevier Ltd All rights reserved DOI 10.1016/j.str.2006.01.007

Crystal Structure of the HP1-EMSY Complex Reveals an Unusual Mode of HP1 Binding

Ying Huang,[1] Michael P. Myers,[1] and Rui-Ming Xu[1,*]
[1] Cold Spring Harbor Laboratory
Cold Spring Harbor, New York 11724

Summary

Heterochromatin protein-1 (HP1) plays an essential role in both the assembly of higher-order chromatin structure and epigenetic inheritance. The C-terminal chromo shadow domain (CSD) of HP1 is responsible for homodimerization and interaction with a number of chromatin-associated nonhistone proteins, including EMSY, which is a BRCA2-interacting protein that has been implicated in the development of breast and ovarian cancer. We have determined the crystal structure of the HP1β CSD in complex with the N-terminal domain of EMSY at 1.8 Å resolution. Surprisingly, the structure reveals that EMSY is bound by two HP1 CSD homodimers, and the binding sequences differ from the consensus HP1 binding motif PXVXL. This structural information expands our understanding of HP1 binding specificity and provides insights into interactions between HP1 homodimers that are likely to be important for heterochromatin formation.

Introduction

Heterochromatin protein-1 (HP1) is a conserved chromatin-associated protein important for both the assembly of higher-order chromatin structure and regulation of gene transcription (Eissenberg and Elgin, 2000; Maison and Almouzni, 2004). HP1 was first discovered in *Drosophila*, and it plays essential roles in position effect variegation, which is a phenomenon of epigenetic gene silencing through heterochromatin formation (Eissenberg et al., 1990; James et al., 1989; Wustmann et al., 1989). HP1 achieves its silencing function by directly binding to chromatin and spreading across the entire silenced chromosomal region (Maison and Almouzni, 2004). Similar mechanisms of epigenetic control of gene expression are employed in organisms from yeasts to humans and in diverse biological processes ranging from mating-type switching to X chromosome inactivation (Grewal and Elgin, 2002; Jaenisch and Bird, 2003; Nusinow and Panning, 2005; Rusche et al., 2003). HP1 and localized heterochromatic foci have also been implicated in cellular senescence and cancer development (Narita et al., 2003; Nielsen et al., 2001).

In *Drosophila* and mammals, there are three HP1 homologs: HP1α, HP1β, and HP1γ. The three proteins are highly related, yet they are distributed in distinct chromosomal regions. HP1α and HP1β are predominantly associated with heterochromatin, while HP1γ is mostly euchromatic (Maison and Almouzni, 2004). All HP1 proteins have a highly conserved chromo domain at the N terminus and a well-conserved chromo shadow domain (CSD) at the C terminus (Aasland and Stewart, 1995; Eissenberg, 2001; Paro and Hogness, 1991). The chromo domain of HP1 specifically recognizes the N-terminal tail of histone H3 methylated at lysine-9, which provides a structural basis in support of the "histone code" hypothesis (Bannister et al., 2001; Jacobs and Khorasanizadeh, 2002; Jenuwein and Allis, 2001; Lachner et al., 2001; Nielsen et al., 2002). The CSD is responsible for HP1 homodimerization and interaction with a number of nonhistone nuclear proteins (Brasher et al., 2000; Cowieson et al., 2000; Li et al., 2002; Ye et al., 1997). It has been demonstrated that the HP1 CSD and the hinge region are important for targeting HP1 proteins to specific genomic loci (Smothers and Henikoff, 2001).

Various proteins important for chromatin structure and gene transcription have been found to interact with the HP1 CSD (Lechner et al., 2005; Li et al., 2002). They include the large subunit of chromatin assembly factor-1 (CAF-1) (Brasher et al., 2000; Lechner et al., 2000; Murzina et al., 1999) and components of the cellular transcriptional machinery (Vassallo and Tanese, 2002). In addition, the largest subunit of the origin recognition complex (Orc1) interacts with full-length HP1 rather than with individual domains (Lidonnici et al., 2004; Pak et al., 1997). In the budding yeast *Saccharomyces cerevisiae*, the interaction between the silencer bound Orc1 with silencing information regulator-1 (Sir1) is important for the establishment of the silenced chromatin domain at the cryptic mating-type loci (Gardner et al., 1999; Triolo and Sternglanz, 1996; Zhang et al., 2002). In higher eukaryotes, the interaction between Orc1 and HP1 is likely to play an analogous role in the formation of heterochromatin (Shareef et al., 2003).

A recent addition to the list of HP1-interacting proteins is EMSY, which was identified as a protein that interacts with the exon 3-encoded transcriptional activation domain of breast cancer susceptibility gene 2 (BRCA2) (Hughes-Davies et al., 2003). EMSY has been implicated in BRCA2-associated sporadic breast and ovarian cancer, but an understanding of its cellular functions is lacking. EMSY is a large protein containing an N-terminal domain (ENT) found in many plant proteins, and the ENT domain is followed by an amino acid sequence involved in binding HP1 and BS69, the latter of which was initially isolated as an adenoviral E1A binding protein that recognizes the PXLXP sequence motif (Ansieau and Leutz, 2002; Hateboer et al., 1995; Hughes-Davies et al., 2003).

The EMSY sequence implicated in HP1 binding, RLVPL, deviates from the consensus HP1 binding sequence PXVXL (positions −2 to +2) at the −2 position (Smothers and Henikoff, 2000; Thiru et al., 2004). In the structure of the HP1 CSD-CAF-1 peptide complex, the proline at position −2 occupies a shallow hydrophobic pocket (Thiru et al., 2004) that an arginine would not be able to bind. To determine the exact mode of interaction between the HP1 CSD and EMSY, we solved a 1.8 Å resolution cocrystal structure of the HP1β CSD in complex with an N-terminal EMSY fragment encompassing the ENT domain and the HP1/BS69 binding sites (EMSY-N,

*Correspondence: xur@cshl.edu

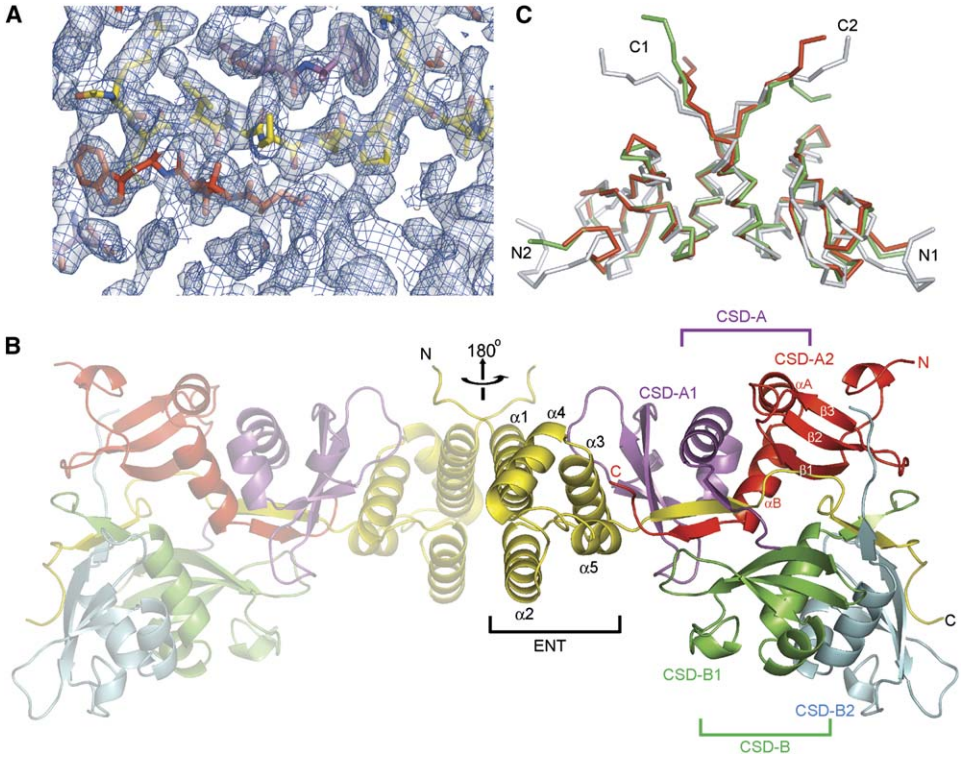

Figure 1. Structure Determination

(A) A section of the MAD-phased electron density map surrounding the ENT-proximal HP1 binding site. The 2.3 Å map is shown in a filled chicken-wire representation and is contoured at the 1.0σ level. The refined model of the EMSY-HP1 CSD complex (stick model) is superimposed onto the map. EMSY, HP1 CSD-A1, and HP1 CSD-A2 molecules are shown with carbon bonds colored yellow, magenta, and red, respectively.

(B) The overall structure of the EMSY-HP1 CSD complex shown in a ribbon representation. One asymmetric unit contains half of the EMSY homodimer, and the other half (shaded) is related by a 2-fold symmetry. EMSY is colored yellow, the ENT-proximal CSD dimer (CSD-A) is shown in magenta (CSD-A1) and red (CSD-A2), and the ENT-distal dimer (CSD-B) is shown in green (CSD-B1) and cyan (CSD-B2).

(C) Superposition of the backbones of CSD-A (red), CSD-B (green), and the HP1 CSD dimer of the HP1-CAF-1 complex (white).

aa 9–139). The structure reveals an unexpected binding mode for HP1 CSD to EMSY, which broadens our understanding of HP1 binding specificity, and the structural information should be useful for identifying new HP1 binding proteins.

Results

Overall Structure

The structure was solved by the method of multiwavelength anomalous dispersion (MAD) by using data collected from a crystal with seleno-methionine (SeMet)-substituted HP1 CSD. The high-quality MAD-phased electron density allows for unambiguous tracing of main chains and assignment of amino acids side chains (Figure 1A). The modeled EMSY-N and the HP1 CSD complex contains two HP1 CSD homodimers (aa 109–175) bound to one EMSY-N (aa 9–124) in the asymmetric unit (Figure 1B). The 1.8 Å refined model has excellent stereochemical quality: 90.3% of the residues are in the most-favored region of the Ramachandran plot, and none are in the disallowed region, assessed by the PROCHECK program (Laskowski et al., 1993). Detailed statistics of structure determination and refinement are shown in Table 1.

The ordered EMSY-N structure contains the ENT domain (aa 9–97) and a C-terminal extension containing the HP1 binding sites (aa 98–124). The structure of the ENT domain has been previously described (Chavali et al., 2005). It consists mainly of five α helices: α1 and α2 are antiparallel to each other, α3 and α5 are packed on the same side of α1 and α2, and α4 contacts the C-terminal end of α1 (Figure 1B). As noted in the work of Chavali et al. (2005), the ENT domain forms a homodimer, and the two ENT monomers are related by a 2-fold crystallographic symmetry in our structure (Figure 1B). Helices α1, α2, and α4 are engaged in extensive dimeric interactions. The dimeric interface buries a total of ~2976 $Å^2$ of pairwise solvent-accessible area, calculated with a probe radius of 1.4 Å. The C-terminal extension of EMSY-N, which harbors the HP1 binding sites, projects away from the dimeric interface (Figure 1B).

The HP1 CSD has a mixed α/β structure (Brasher et al., 2000; Cowieson et al., 2000): three β strands (β1–β3) at the N-terminal end form an antiparallel sheet, which is stabilized by the packing of two approximately perpendicularly oriented α helices (αA and αB) on one side of the β sheet. The two helices are responsible for homodimer formation, and they interact with the same helices of another CSD monomer (Figure 1B). The two HP1 CSD homodimers in the structure, denoted CSD-A and CSD-B, have very similar monomeric structures. In addition, the two monomers use similar modes to pack against each other. The two homodimers can be superimposed

Table 1. Summary of Crystallographic Analysis

Data Sets	Native	Se (Inflection)	Se (Peak)	Se (Remote)
Resolution (Å)	1.8	2.21	2.25	2.25
Measured reflections	303,912	219,016	210,023	213,824
Unique reflections	55,874	58,473	55,243	55,747
Average I/σ	15.8	22.0	21.8	20.4
Completeness (%)	97.0 (96.0)	98.7 (100)	98.5 (100)	99.0 (100)
R_{merge} (%)[a]	5.2 (22.0)	5.8 (9.6)	6.3 (8.8)	6.4 (10.3)
Phasing power[b]		2.41	2.65	1.84
Figure of merit	0.71 (total)	0.43	0.43	0.33
Refinement				
Resolution range (Å)	50–1.8			
R factor/R_{free} (%)[c]	21.9/24.3			
Protein atoms (number)	3,059			
Solvent molecules				
Water	386			
Sulfate	4			
Rms deviations				
Bond lengths (Å)	0.004			
Bond angles (°)	1.197			

[a] $R_{merge} = \Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$, where I and <I> are the averaged intensity of multiple measurements of the same reflection. The summation is over all of the observed reflections.
[b] Phasing power = rms ($\langle F_H\rangle/E$), where F_H is the calculated structure factor of the heavy atoms, and E is the residual lack of closure.
[c] R factor = $\Sigma||F_o| - |F_c||/\Sigma|F_o|$, where F_o denotes the observed structure factor amplitude, and F_c denotes the structure factor calculated from the model.

with a root-mean-square (rms) deviation of 0.56 Å, using the α/β core residues (aa 118–168) from both monomers for alignment (Figure 1C). Using the same set of Cα atoms, the NMR structure of the HP1β CSD (PDB ID: 1S4Z) (Thiru et al., 2004) dimer superimposes with CSD-A, which is the homodimer bound to the ENT-proximal site, with an rms deviation of 2.2 Å (Figure 1C). Principal differences lie at the loops connecting β1-β2 and β2-β3, as well as at both the N and C termini (Figure 1C). The conformational differences are likely due to their involvement in protein-protein interactions with EMSY or with the other HP1 dimer, which will be described in detail below. Analogous to the binding of the CAF-1 peptide (Thiru et al., 2004), the C-terminal end of EMSY binds HP1 CSD homodimers mainly through the C-terminal tails of the HP1 CSD.

HP1 Binding to the ENT-Proximal Site

The first HP1 CSD homodimer, CSD-A, binds to a region adjacent to the helical ENT domain, and it also makes a number of direct contacts with the globular ENT domain. The two monomers of CSD-A, termed CSD-A1 and CSD-A2, bind to EMSY asymmetrically: CSD-A1 is closer to the ENT domain, while CSD-A2 is farther away (Figure 1B). The CSD-A homodimer interacts mainly with a consecutive stretch of EMSY residues (aa 97–106) C-terminal to the ENT domain, which is in an extended conformation and sandwiched between the C termini of two CSD monomers. CSD-A also contacts the ENT domain directly: CSD-A1 interacts with the globular ENT domain through residues located on the β sheet, and CSD-A2 contacts the ENT domain via its C terminus (Figure 1B).

Residues 98–102 of EMSY, with the sequence of Arg-Leu-Val-Pro-Leu, form an antiparallel intermolecular β sheet pairing with residues 167–171 of CSD-A2. The EMSY residues 101–104 form a parallel intermolecular paring with residues 168–171 of CSD-A1 (Figure 1B). Pro101 of EMSY occupies a central position within the symmetric CSD-A homodimer, and its side chain is situated in a symmetric hydrophobic pocket formed by Tyr164 and Leu168 of both HP1 CSD monomers (Figures 2A and 2B). We call this binding site position 0, following the nomenclature used in the description of the HP1β CSD-CAF-1 peptide complex structure, in which a valine occupies the corresponding site (Thiru et al., 2004). As in the structure with the bound CAF-1 peptide, positions −2 and +2 are additional major interacting sites between the HP1 CSD and EMSY (Figure 2B). Leu99 of EMSY binds at the −2 position, which is formed with Leu168 and Trp170 of CSD-A2 and Ala125, Leu132, and Asp127 of CSD-A1 (Figure 2A). In addition, the main chains of Thr169 of CSD-A2 and Thr126 of CSD-A1 form part of the binding pocket. Met103 of EMSY occupies the +2 site, which is formed with Leu168 and Trp170 of CSD-A1 and Ala125, Phe163, and Arg167 of CSD-A2 (Figure 2A). The composition of the −2 and +2 positions clearly indicates the asymmetry of the two binding sites.

The main chain atoms of the EMSY residues at positions −1, +1, −3, and +3 are involved in intermolecular β pairing with both molecules of CSD-A, but their side chains point away from the center of the CSD dimer and are only involved in mostly van der Waals contacts with the CSD homodimer (Figures 2A and 2B). Of note is the binding of EMSY Arg98 at position −3. It is sandwiched between Arg167 of CSD-A1 and Tyr173 of CSD-A2 (Figure 2A). The small size of Pro104 at the +3 position is probably needed to accommodate the binding of a second CSD dimer, CSD-B, at the distal site. Arg105 of EMSY is at the +4 position; the aliphatic portion of the side chain packs against Trp170 of CSD-A1, and its Nε atom makes a hydrogen bond with the Oδ1 atom of Asp127 from CSD-A2 (Figure 2A). Finally, Leu106 of EMSY at position +5 contacts Thr126, Leu135, and Leu146, all from CSD-A2, through hydrophobic interactions.

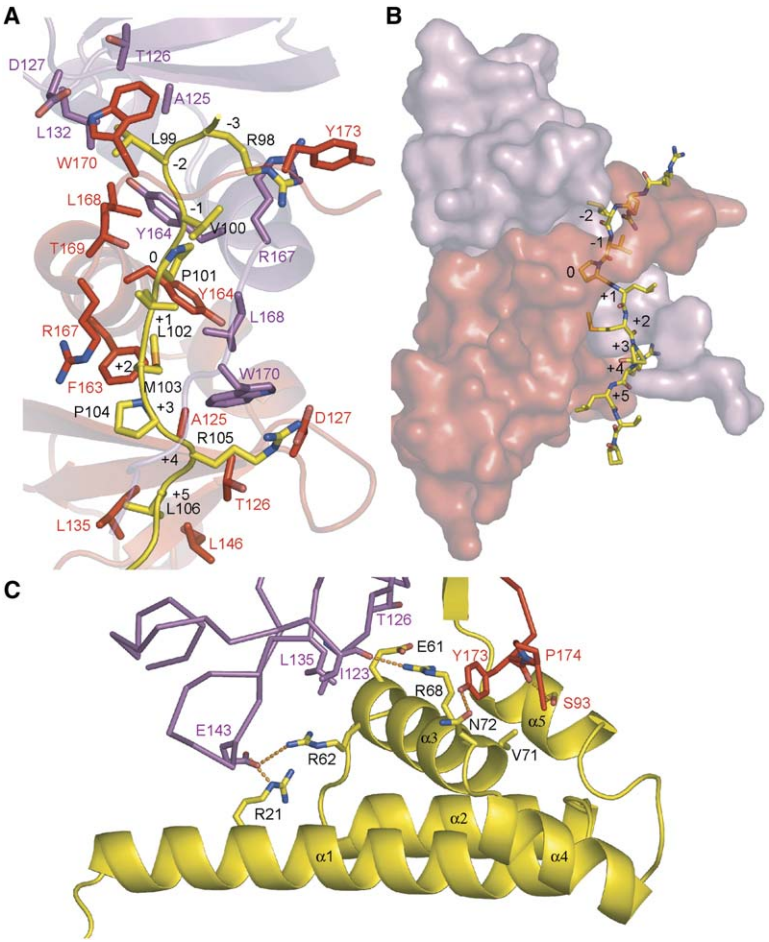

Figure 2. The ENT-Proximal Site

(A) The binding of CSD-A to the ENT-proximal site. The main chain of the ENT-proximal sequence is shown as a coil, and CSD-A is shown in a ribbon representation (with the C-terminal short β strands in coils for viewing clarity). The residues involved in interactions between EMSY and CSD-A are shown in stick models; the same coloring scheme as that in Figure 1 is used.

(B) The CSD-A homodimer is shown in a surface representation, and the ENT-proximal sequence is shown in a stick model. The binding positions are labeled.

(C) Interactions between the CSD-A homodimer and the globular ENT domain. Hydrogen bonds are indicated with dashed lines.

CSD-A makes limited contacts with the globular ENT domain through the C terminus of CSD-A2 and β sheet residues of CSD-A1 (Figure 1B). Tyr173 of CSD-A2 makes van der Waals contacts with Arg68, Val71, and Asn72 and hydrogen bonds to Oδ1 of Asn72, all located on α3 of the ENT domain (Figure 2C). In addition, the main chain atoms of Tyr173 and Pro174 of CSD-A2 make hydrogen bonds with the main chain and side chain oxygen atoms of Ser93, located in α5 of EMSY. The β sheet surface of CSD-A1 is close to the ENT domain surface region formed by the N-terminal ends of helices α1 and α3. Glu143 located in the β2-β3 turn of CSD-A1 hydrogen bonds with Arg21 in α1 and Arg62 in α3 of the ENT domain (Figure 2C). Leu135 in β2 of CSD-A1 makes van der Waals contacts with Glu61 and Arg62 in α3 of the ENT domain. Furthermore, the carbonyl of Ile123 in β2 hydrogen bonds with Arg68 in α3, and Thr126 makes van der Waals contacts with Glu61, also in α3 of the ENT domain. These contacts appear to augment the binding of CSD-A to the previously described ENT-proximal site.

HP1 Binding to the ENT-Distal Site

The second HP1 CSD homodimer, CSD-B, interacts with residues 106–124 of EMSY (Figures 1B and 3). Because of the close juxtaposition of the two HP1 CSD homodimers, Leu106 and Thr110 of EMSY interact with both CSD-A and CSD-B: Leu106 of EMSY binds in a pocket formed by Thr126, Leu135, and Leu146 of CSD-A2 and Tyr173 of CSD-B2, while Thr110 contacts Tyr173 of CSD-A1 and Thr126 of CSD-B1 (Figure 4B).

As in the ENT-proximal HP1 binding site, EMSY residues at positions −3 to +1 (aa 111–115) are involved in antiparallel intermolecular β pairing with residues 168–171 of CSD-B2. EMSY residues at positions 0 to +2 form a parallel intermolecular β pairing with residues 169–171 of CSD-B1 (Figure 1B). Val114 of EMSY binds in the hydrophobic pocket at position 0 of CSD-B, in a manner similar to the binding of Pro101 of EMSY at the corresponding CSD-A location (Figure 3). Phe112 of EMSY at position −2 interacts extensively with residues forming the −2 binding site of CSD-B, which consist of CSD-B1 residues Ala125, Phe163, Tyr164, and the aliphatic portion of Arg167, and CSD-B2 residues Leu168 and Trp170 (Figure 3A). Instead of a bulkier aliphatic residue or methionine, as in the ENT-proximal site and HP1 binding sequences in previous studies, an alanine, Ala116, of EMSY occupies position +2 of the ENT-distal site. The threonines at the −1 and +1 positions of EMSY, Thr113, and Thr115 are engaged in similar interactions with CSD-B: contacting Thr169 of both CSDs and His171 of one of the molecules in the CSD-B homodimer (Figure 3A). Ala118 of EMSY at the +4 position interacts with Trp170 and Pro174 of CSD-B1 and Thr126 of CSD-B2 through van der Waals contacts, while EMSY C-terminal residues Val119 and Val124

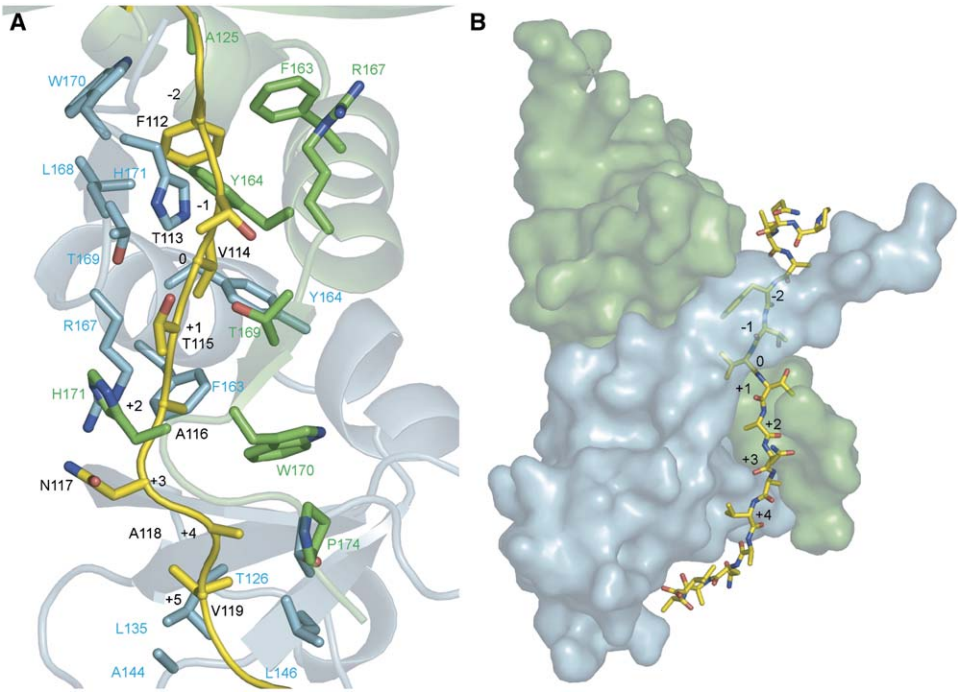

Figure 3. The ENT-Distal Site

(A) The binding of the CSD-B dimer to the ENT-distal site. The residues involved in the interactions are shown in a stick model.

(B) A surface representation showing the binding of CSD-B to the ENT-distal sequence, which is shown in a stick model. The same coloring scheme as that in Figure 1B (CSD-B1 in green, CSD-B2 in cyan, and EMSY in yellow) is used.

exclusively interact with CSD-B2 residues located on the β sheet surface, which include Thr126, Leu135, Ala144, and Leu146 (Figure 3A).

HP1-HP1 Interactions

We noted earlier that the two CSD homodimers are in close juxtaposition, and they interact with each other. Extensive interactions are observed between CSD-A2 and CSD-B1, mainly between residues in the C-terminal end of one protein and residues located at the N-terminal end of β1 and the β2-β2 turn of another. Many of the interactions are hydrogen bonds between the side chains of charged residues in one CSD dimer and main chain atoms of another CSD dimer. It is interesting to note that the side chain-main chain hydrogen bonds between the two CSD proteins mimic a pattern of β sheet hydrogen bonding (Figure 4A). The main chain-side chain β sheet-type pairing occurs between two adjacent CSD molecules, CSD-A2 and CSD-B1, belonging to two different CSD homodimers. The interactions are approximately symmetric, i.e., the charged side chains in CSD-B1 make hydrogen bonds with main chain atoms of C-terminal residues in CSD-A2, and, conversely, the charged side chains in CSD-A2 make hydrogen bonds with main chain atoms of C-terminal residues in CSD-B1 (Figure 4A).

Additional interactions between the two CSD dimers involve the packing of the C-terminal tail of CSD-A1 (aa 171–173) against the β sheet surface of CSD-B1 and the corresponding packing of the C-terminal tail of CSD-B2 and the β sheet surface of CSD-A2. In particular, His171 of CSD-A1 makes a hydrogen bond with the amide group of Ala144 of CSD-B1, and Tyr173 of CSD-A1 binds in a position between Thr126 and Leu135 and makes a hydrogen bond with the amide group of Ala125 of CSD-B1 (Figure 4B). The corresponding residues at the C-terminal end of CSD-B2 interact with CSD-A2 differently, likely due to overlapping of two HP1 binding sites in EMSY.

Comparison of HP1 CSD Bindings

Superposition of the ENT-proximal, ENT-distal, and CAF-1 bound HP1 CSD dimers reveals that the three CSD bound sequences are reasonably well aligned in a region encompassing residues from positions −1 to +5, with the largest deviation of Cα positions about 2.3 Å. Larger deviations are observed for amino acids outside of this region, possibly due to their contacts with the more flexible C-terminal ends of HP1 CSD (Figure 5A). Previous studies have revealed a HP1 CSD consensus binding motif, PXVXL, where the proline at position −2 and the valine at position 0 are crucial for HP1 binding (Smothers and Henikoff, 2000; Thiru et al., 2004). The structural alignment reveals that the corresponding residues are LVPLM and FTVTA at the ENT proximal and distal sites, respectively (Figure 5A). In particular, the proline at position −2 is replaced by a leucine in the ENT-proximal site and a phenylalanine in the ENT-distal site, while the valine at position 0 can be a proline.

An examination of the structure reveals that the proline at position 0 binds in a manner similar to the valine in the HP1-CAF-1 structure. Both residues are similar in size, and they bind in the symmetric hydrophobic

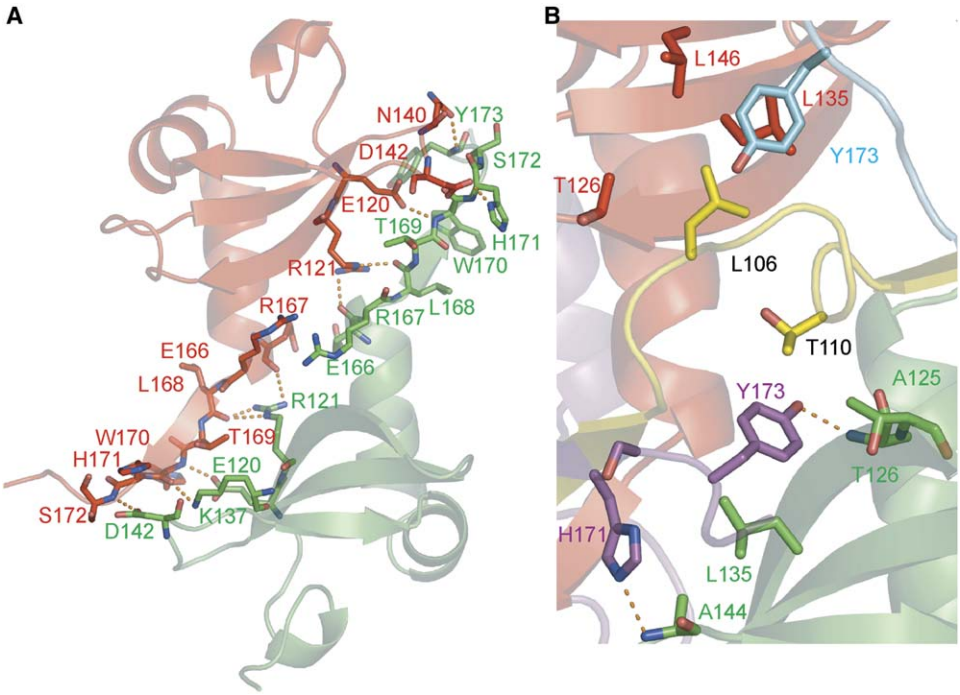

Figure 4. Interactions between HP1 CSD Dimers

(A) An extensive network of side chain-main chain hydrogen bonds between CSD-A2 (red) and CSD-B1 (green) is shown; dashed lines indicate the hydrogen bonds.

(B) Packing of the C-terminal residues of CSD-A1 (magenta) with CSD-B1 (green). Also shown are EMSY residues contacting both CSD dimers. The same coloring scheme as that in Figure 1B is used.

pocket of the HP1 CSD homodimer. The binding site for the residue at position −2 appears to be more dynamic. In the HP1-CAF-1 structure, the proline binds equidistantly from Phe163 located on αB of one HP1 CSD monomer and Trp170 at the C terminus of another monomer. In the present structure, the C termini of the HP1 CSD are farther away from helix 2, which results in preferential packing of Leu99 in the proximal site and Phe112 of the distal site with Trp170 (Figure 5A). This observation demonstrates that the position −2 binding site is capable of accommodating a variety of hydrophobic residues, owing to the flexibility of the HP1 C terminus.

It was unexpected that two HP1 CSD homodimers would be bound to one EMSY monomer. Because the two homodimers contact each other and several EMSY residues are contacted by both homodimers, we tested whether the two HP1 CSD homodimers bind to EMSY independently. We generated two EMSY-N variants, one lacking the second HP1 binding site (aa 9–110), termed EMSY-N1, and another with the first binding sites (aa 97–109) removed, termed EMSY-N2. Figure 5B shows that GST-fused HP1 CSD efficiently pulls down EMSY-N1 (lane 9), but not EMSY-N2 (lane 11), which suggests that the ENT-proximal site is an independent HP1 binding site, while the distal site is not. To determine whether certain amino acids in the proximal region are required for the binding of HP1 to the distal sites, we generated a P101Q mutation in EMSY-N1 (lane 3) and EMSY-N (lane 5), which changes the proline at position 0 to a glutamine. As shown in Figure 5B, the P101Q mutant of EMSY-N1 lost its ability to bind to HP1 CSD (lane 10); furthermore, we cannot detect the binding of HP1 CSD to the P101Q mutant of EMSY-N (lane 12), which has an intact distal HP1 binding site. Conversely, if Val114 at position 0 of the distal site is mutated to a glutamate but the proximal site remains intact, the binding of the HP1 CSD to EMSY-N is little affected (lane 13). Thus, we conclude that the ENT proximal site is an independent HP1 binding site but that the distal site is not.

To verify the 1:4 EMSY-N:HP1 CSD binding stoichiometry, we performed a stable isotope mass spectrometry quantitation experiment. HP1 was mixed with EMSY or with EMSY-N1 (which can only bind one HP1 CSD homodimer) in excess, and the resulting mixtures were separated on a Superdex-200 gel filtration column to isolate the complex from unbound components (Figures 6A and 6B). The HP1 CSD-EMSY-N (complex A) and HP1 CSD-EMSY-N1 (complex B) complexes were then denatured and digested with trypsin. The complex A digest was acetylated with acetic anhydride, while the complex B peptides were acetylated with deuterated acetic anhydride. This acetylation strategy results in a mass difference of 3 Da per incorporated acetyl group between two identical peptides. The two digests were mixed together and analyzed by MALDI-TOF mass spectrometry. The areas of three peptides coming from EMSY-N and EMSY-N1 were used to normalize the amount of two complexes (a representative spectrum is shown in Figure 6C), and the normalized ratios between the areas of the same peptide bound to EMSY-N and EMSY-N1 reveal that EMSY-N binds approximately two times more moles of HP1 than EMSY-N1 (2.1 ± 0.1 averaged with three different peptides). Thus, we conclude that the 1:4 EMSY-N:HP1 CSD stoichiometry observed in the

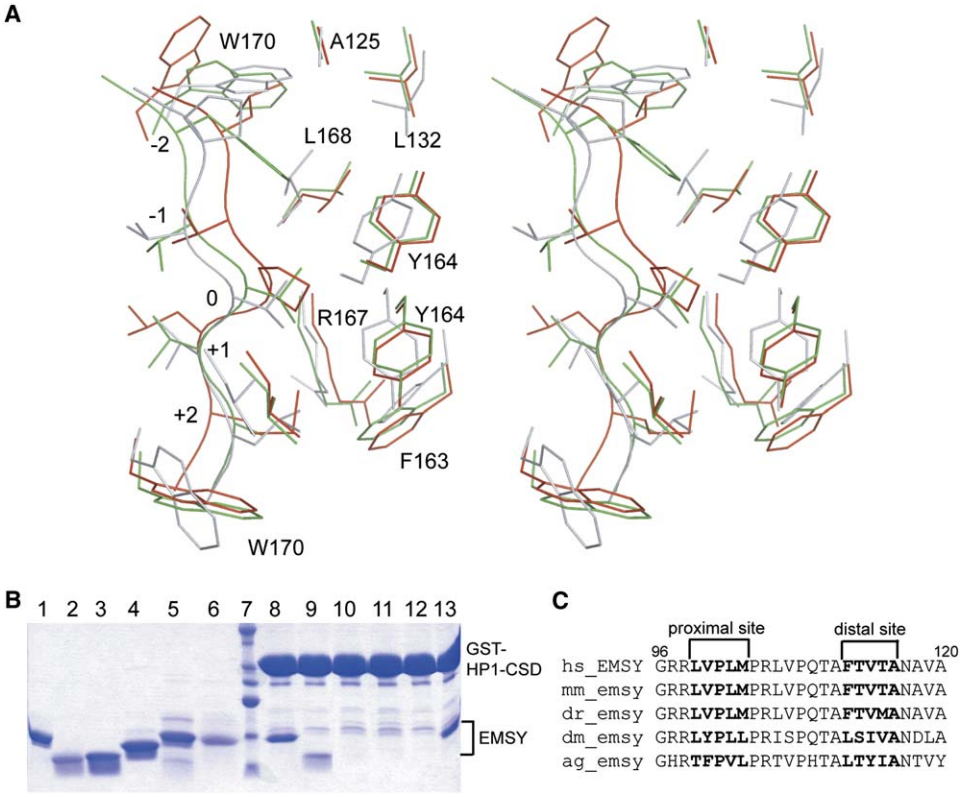

Figure 5. Comparison of HP1 CSD Bindings

(A) A stereodiagram showing the superposition of CSD-A (red), CSD-B (green), and the CSD dimer of the HP1-CAF-1 complex (white) bound to their cognate binding sequences.

(B) Coomassie-stained gel showing the binding of the HP1 CSD to EMSY-N and its variants. Lanes 1–6, input samples of partially purified EMSY-N, EMSY-N1, EMSY-N1(P101Q), EMSY-N2, EMSY-N(P101Q), and EMSY-N(V114E), respectively; lane 7, molecular weight marker; lanes 8–13, proteins corresponding to those shown in lanes 1–6 bound to GSH resins preincubated with GST-HP1β CSD.

(C) Alignment of the HP1 binding sequences of EMSY orthologs. Shown from the top to the bottom are the sequences of human, mouse, zebrafish, *Drosophila*, and mosquito. The two binding sites are shown in bold face.

crystal structure also represents a binding stoichiometry in solution.

Discussion

HP1 proteins play diverse roles in higher chromatin structure and gene regulation, and their various interaction partners have been described. The C-terminal CSD is essential for most of the protein-protein interactions studied to date. Previous structural and biochemical studies have defined a short consensus HP1 CSD binding motif, PXVXL, which has greatly facilitated the understanding of the structural basis for HP1 binding specificity (Smothers and Henikoff, 2000; Thiru et al., 2004). This study reveals that the HP1 CSD has considerable flexibility in accommodating the binding of different amino acid sequences. Instead of the PXVXL motif, the valine at position 0 can be a proline, and the proline at position −2 can be a hydrophobic residue such as a leucine or a phenylalanine. Thus, the HP1 binding motif can be generalized to ΦX(V/P)X(L/M/V), where Φ stands for hydrophobic residues. As shown earlier, we found that the ENT-distal site is not capable of binding HP1 CSD independently. This is probably due to the presence of an alanine at position +2, which, in comparison to the bulkier aliphatic residues or methionine, loses the benefit of tighter interaction in the hydrophobic binding pocket. Our structural studies should facilitate the identification of new HP1-interacting proteins and provide mechanistic insights into the broad role of HP1 protein-protein interactions.

A surprising finding of our study is that one EMSY monomer binds to two HP1 CSD homodimers. In agreement with previous binding studies, a shorter EMSY-N fragment encompassing residues 1–108 binds one HP1 CSD dimer, although the observed binding mode in our crystal structure differs from that modeled according to the structure of HP1 CSD in complex with the CAF-1 peptide (Ekblad et al., 2005). Based on an unpublished observation, it has been suggested that a longer fragment of EMSY, encompassing residues 1–140, which contains both HP1 binding sites observed in our structure, also bind to HP1 CSD with same stoichiometry (Ekblad et al., 2005). The binding data are not necessarily incompatible with our structural and mass spectrometry results. This is because, first, determination of binding stoichiometry by isothermal titration calorimetry (ITC) requires the input of a binding model, which, in the absence of the knowledge about the second binding site, may lead to a different interpretation of the ITC results; second, our GST pull-down experiments demonstrated that the second binding site does not function

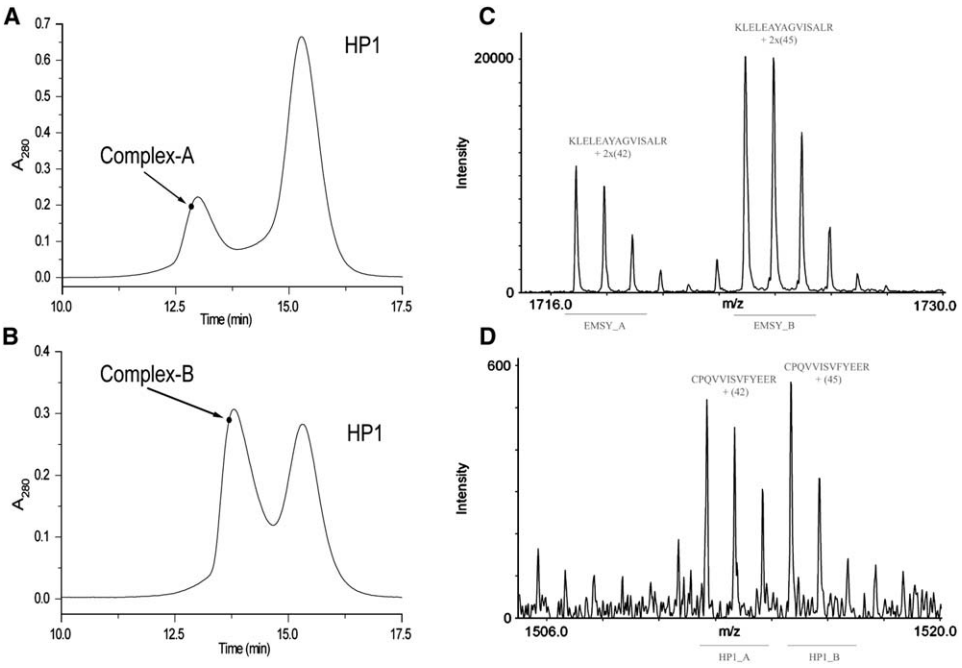

Figure 6. Binding Stoichiometry in Solution

(A) A Superdex-200 column chromatogram showing the separation of the HP1 CSD-EMSY-N complex (complex A) from unbounded HP1 CSD. The fraction used for mass spectrometry analysis is indicated with an arrow.

(B) Separation of the HP1 CSD-EMSY-N1 complex (complex B) from unbounded HP1 CSD on a Superdex-200 column. The arrow points to the fraction used for mass spectrometry analysis.

(C) A representative spectrum of an acetylated EMSY peptide from EMSY-N (EMSY-A) and the same peptide from EMSY-N1 with a deuterated acetyl group (EMSY-B). The sequence of the peptide is shown above the spectrum.

(D) A spectrum of a representative acetylated HP1 CSD peptide from the same complex A and complex B (deuterated) mixture as in (C). The peptide sequence is indicated.

independently, which suggests that, under certain conditions, the EMSY-HP1 complex may exist as an equilibrium of subcomplexes in solution with the first or both HP1 binding sites of EMSY occupied. It is possible that interactions between HP1 CSD dimers stabilize the binding of the second HP1 CSD dimer to the ENT-distal site. An examination of EMSY orthologs from different species shows that both binding sites are highly conserved, with the exception that the distal sites of the *Drosophila* and mosquito proteins have an isoleucine and a tyrosine at position 0, respectively (Figure 5C). We do not know if HP1 can also bind to these two proteins in a manner similar to what we observed here. It is possible that the binding of a second HP1 homodimer to EMSY is specific to certain species.

It is interesting that the close proximity of the two HP1 binding sites allows two adjacent CSD dimers to interact with each other. To our knowledge, this represents the first observation of interactions between HP1 CSD homodimers. The structure shows that the two CSD homodimers have good structural and chemical complementarities, which suggests that, under favorable conditions, HP1 CSD homodimers could form higher-order structures. The result may be of great biological interest because the HP1 dimer is known to spread a long distance, yet no interaction between HP1 dimers has been described. It is possible that interactions between HP1 homodimers in the vicinity of nucleosome arrays can facilitate efficient polymerization and spreading of HP1, which is the hallmark of HP1 function in heterochromatin formation.

Experimental Procedures

Protein Purification and Crystallization

A cDNA fragment encoding residues 9–139 of EMSY (EMSY-N) was cloned into a modified pGEX-KG vector bearing a PreScission protease (Amersham) cleavage site C-terminal to the GST tag. GST-tagged EMSY-N was expressed by using the BL21(DE3) strain of *E. coli* at 18°C overnight. The GST fusion protein was first purified on a glutathione-agarose (GSH, Sigma) column, followed by on-column digestion with the PreScission protease. The eluted protein was further purified by gel-filtration column chromatography. Human HP1β CSD (aa 104–175) was expressed as a poly(histidine)-tagged protein by using the pET28a vector (Novagen). The his-tagged protein was first purified on a Ni-chelating column. After cleavage of the His tag with thrombin, HP1β CSD was further purified by gel-filtration column chromatography. To obtain the HP1β CSD-EMSY-N complex, purified EMSY-N and HP1β CSD were mixed and incubated on ice for 1 hr. To allow for efficient formation of the complex, HP1β CSD was added in excess. The HP1β CSD-EMSY-N complex was purified on a gel-filtration column (in a buffer containing 10 mM Tris [pH 8.0], 0.1 M NaCl, 1 mM DTT), which separated the complex from the excessive HP1β CSD homodimer. The HP1β CSD-EMSY-N complex was crystallized by using the hanging drop vapor diffusion method at a protein concentration of ~20 mg/ml. The well buffer contains 0.1 M Tris (pH 8.0) and 1.8 M ammonium sulfate. Crystals grew over a period of 2–3 days, with typical dimensions of 0.3 × 0.3 × 0.3 mm^3. The crystal belongs to the $P2_12_12$ space group with cell dimensions of a = 143.4, b = 63.87, c = 66.64 Å, and it contains one EMSY-N and four HP1β CSD molecules per asymmetric unit.

Data Collection and Structure Determination

X-ray diffraction data were collected at the X26C beamline of the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory (BNL). Both the native and seleno-methionine (SeMet)-substituted crystals were flash frozen under a cold nitrogen stream (100 K) during data collection. The data were processed by using the HKL program suite (Otwinowski and Minor, 1997). Initial phases were determined by using the SeMet multiwavelength anomalous diffraction (MAD) data sets collected from a crystal of SeMet in HP1β CSD. The SOLVE and CNS programs were used for location of Se positions, phasing, and density modification (Brunger et al., 1998; Terwilliger and Berendzen, 1999). The graphics program O was used for model building (Jones et al., 1991), and refinement was carried out by using CNS (Brunger et al., 1998). Phasing and refinement statistics are listed in Table 1. Figures were prepared with PyMOL (DeLano, 2002).

Mutagenesis, GST Pull-Down, and Mass Spectrometry Experiments

Deletion and amino acid substitution mutants of EMSY-N were generated by PCR, and the recombinant proteins were prepared in a procedure similar to that described for the wild-type protein used for crystallization. For pull-down experiments, GST-fused HP1β CSD was expressed in *E. coli*. The lysate supernatant was incubated with GSH resins for 30 min on ice, which was followed by washing the resins with a buffer containing 20 mM Tris (pH 8.0), 0.5 M NaCl, 1 mM EDTA, 1 mM DTT, 0.02% NP-40, and 5% glycerol. The washed resins were divided into several tubes, each containing ~20 μg proteins as determined by the Bradford assay. Approximately 50 μg purified EMSY-N or its variants were mixed with the GST-HP1 CSD bound resins, and the mixture was incubated on ice for 1 hr. The resins were then washed with the washing buffer described above, except that the NaCl concentration had been reduced to 0.1 M. The proteins bound to the resins were then analyzed by SDS-PAGE.

Gel-filtration fractions were supplemented with acetonitrile to 10% and TCEP to 10 mM and were boiled for 5 min. Trypsin (20 ng) was added to the cooled reaction and incubated at 37°C for 12 hr. The digested peptides were acetylated by the addition of 20 μl acetic anhydride (complex A) or deuterated acetic anhydride (D_6) (complex B). The acetylation was allowed to proceed at room temperature for 1 hr and was then quenched by the addition of trifluoracetic acid to 0.5%. The two complexes were immediately mixed and desalted by micro reverse-phase columns. The resulting mixture was mixed with MALDI matrix (saturated trans-3-indole acrylic acid in 50% acetonitrile, 0.1% acetic acid) and spotted directly onto the MALDI target. Mass spectra were collected by using an ABI 4700 and were analyzed with Data Explorer (Applied Biosystems) and m/z (Genomic Solutions) software programs. The relative stoichiometry was calculated by using the following formula: $([HP1_A]/[HP1_B]) \div ([EMSY\text{-}N]/[EMSY\text{-}N1])$, where the square bracket represents the area of a peptide peak from the sample whose name is indicated inside the bracket.

Acknowledgments

We thank Annie Heroux and Dieter Schneider for help during data collection at the National Synchrotron Light Source, and we thank Lee Henry for comments on the manuscript. The work was supported in part by the W.M. Keck Foundation and a grant from the National Institutes of Health (GM 63716).

Received: October 19, 2005
Revised: December 9, 2005
Accepted: January 7, 2006
Published: April 11, 2006

References

Aasland, R., and Stewart, A.F. (1995). The chromo shadow domain, a second chromo domain in heterochromatin-binding protein *1*, HP1. Nucleic Acids Res. *23*, 3168–3173.

Ansieau, S., and Leutz, A. (2002). The conserved Mynd domain of BS69 binds cellular and oncoviral proteins through a common PXLXP motif. J. Biol. Chem. *277*, 4906–4910.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.

Brasher, S.V., Smith, B.O., Fogh, R.H., Nietlispach, D., Thiru, A., Nielsen, P.R., Broadhurst, R.W., Ball, L.J., Murzina, N.V., and Laue, E.D. (2000). The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. EMBO J. *19*, 1587–1597.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Chavali, G.B., Ekblad, C.M., Basu, B.P., Brissett, N.C., Veprintsev, D., Hughes-Davies, L., Kouzarides, T., Itzhaki, L.S., and Doherty, A.J. (2005). Crystal structure of the ENT domain of human EMSY. J. Mol. Biol. *350*, 964–973.

Cowieson, N.P., Partridge, J.F., Allshire, R.C., and McLaughlin, P.J. (2000). Dimerization of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. Curr. Biol. *10*, 517–525.

DeLano, W.L. (2002). The PYMOL Molecular Graphics System (San Carlos, CA: DeLano Scientific).

Eissenberg, J.C. (2001). Molecular biology of the chromo domain: an ancient chromatin module comes of age. Gene *275*, 19–29.

Eissenberg, J.C., and Elgin, S.C. (2000). The HP1 protein family: getting a grip on chromatin. Curr. Opin. Genet. Dev. *10*, 204–210.

Eissenberg, J.C., James, T.C., Foster-Hartnett, D.M., Hartnett, T., Ngan, V., and Elgin, S.C. (1990). Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA *87*, 9923–9927.

Ekblad, C.M., Chavali, G.B., Basu, B.P., Freund, S.M., Veprintsev, D., Hughes-Davies, L., Kouzarides, T., Doherty, A.J., and Itzhaki, L.S. (2005). Binding of EMSY to HP1β: implications for recruitment of HP1β and BS69. EMBO Rep. *6*, 675–680.

Gardner, K.A., Rine, J., and Fox, C.A. (1999). A region of the Sir1 protein dedicated to recognition of a silencer and required for interaction with the Orc1 protein in *Saccharomyces cerevisiae*. Genetics *151*, 31–44.

Grewal, S.I., and Elgin, S.C. (2002). Heterochromatin: new possibilities for the inheritance of structure. Curr. Opin. Genet. Dev. *12*, 178–187.

Hateboer, G., Gennissen, A., Ramos, Y.F., Kerkhoven, R.M., Sonntag-Buck, V., Stunnenberg, H.G., and Bernards, R. (1995). BS69, a novel adenovirus E1A-associated protein that inhibits E1A transactivation. EMBO J. *14*, 3159–3169.

Hughes-Davies, L., Huntsman, D., Ruas, M., Fuks, F., Bye, J., Chin, S.F., Milner, J., Brown, L.A., Hsu, F., Gilks, B., et al. (2003). EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. Cell *115*, 523–535.

Jacobs, S.A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science *295*, 2080–2083.

Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. *33* (*Suppl*), 245–254.

James, T.C., Eissenberg, J.C., Craig, C., Dietrich, V., Hobson, A., and Elgin, S.C. (1989). Distribution patterns of HP1, a heterochromatin-associated nonhistone chromosomal protein of *Drosophila*. Eur. J. Cell Biol. *50*, 170–180.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074–1080.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard. (1991). Improved methods for building protein models in electron density

maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to produce both detailed and schematic plots of proteins. J. Appl. Crystallogr. *24*, 946–956.

Lechner, M.S., Begg, G.E., Speicher, D.W., and Rauscher, F.J., 3rd. (2000). Molecular determinants for targeting heterochromatin protein 1-mediated gene silencing: direct chromoshadow domain-KAP-1 corepressor interaction is essential. Mol. Cell. Biol. *20*, 6449–6465.

Lechner, M.S., Schultz, D.C., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd. (2005). The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. Biochem. Biophys. Res. Commun. *331*, 929–937.

Li, Y., Kirschmann, D.A., and Wallrath, L.L. (2002). Does heterochromatin protein 1 always follow code? Proc. Natl. Acad. Sci. USA *99* (*Suppl 4*), 16462–16469.

Lidonnici, M.R., Rossi, R., Paixao, S., Mendoza-Maldonado, R., Paolinelli, R., Arcangeli, C., Giacca, M., Biamonti, G., and Montecucco, A. (2004). Subnuclear distribution of the largest subunit of the human origin recognition complex during the cell cycle. J. Cell Sci. *117*, 5221–5231.

Maison, C., and Almouzni, G. (2004). HP1 and the dynamics of heterochromatin maintenance. Nat. Rev. Mol. Cell Biol. *5*, 296–304.

Murzina, N., Verreault, A., Laue, E., and Stillman, B. (1999). Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor 1 and HP1 proteins. Mol. Cell *4*, 529–540.

Narita, M., Nunez, S., Heard, E., Lin, A.W., Hearn, S.A., Spector, D.L., Hannon, G.J., and Lowe, S.W. (2003). Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. Cell *113*, 703–716.

Nielsen, P.R., Nietlispach, D., Mott, H.R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A.G., Murzina, N.V., and Laue, E.D. (2002). Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature *416*, 103–107.

Nielsen, S.J., Schneider, R., Bauer, U.M., Bannister, A.J., Morrison, A., O'Carroll, D., Firestein, R., Cleary, M., Jenuwein, T., Herrera, R.E., and Kouzarides, T. (2001). Rb targets histone H3 methylation and HP1 to promoters. Nature *412*, 561–565.

Nusinow, D.A., and Panning, B. (2005). Recognition and modification of seX chromosomes. Curr. Opin. Genet. Dev. *15*, 206–213.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Pak, D.T., Pflumm, M., Chesnokov, I., Huang, D.W., Kellum, R., Marr, J., Romanowski, P., and Botchan, M.R. (1997). Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. Cell *91*, 311–323.

Paro, R., and Hogness, D.S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. Proc. Natl. Acad. Sci. USA *88*, 263–267.

Rusche, L.N., Kirchmaier, A.L., and Rine, J. (2003). The establishment, inheritance, and function of silenced chromatin in *Saccharomyces cerevisiae*. Annu. Rev. Biochem. *72*, 481–516.

Shareef, M.M., Badugu, R., and Kellum, R. (2003). HP1/ORC complex and heterochromatin assembly. Genetica *117*, 127–134.

Smothers, J.F., and Henikoff, S. (2000). The HP1 chromo shadow domain binds a consensus peptide pentamer. Curr. Biol. *10*, 27–30.

Smothers, J.F., and Henikoff, S. (2001). The hinge and chromo shadow domain impart distinct targeting of HP1-like proteins. Mol. Cell. Biol. *21*, 2555–2569.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. *55*, 849–861.

Thiru, A., Nietlispach, D., Mott, H.R., Okuwaki, M., Lyon, D., Nielsen, P.R., Hirshberg, M., Verreault, A., Murzina, N.V., and Laue, E.D. (2004). Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. EMBO J. *23*, 489–499.

Triolo, T., and Sternglanz, R. (1996). Role of interactions between the origin recognition complex and SIR1 in transcriptional silencing. Nature *381*, 251–253.

Vassallo, M.F., and Tanese, N. (2002). Isoform-specific interaction of HP1 with human TAFII130. Proc. Natl. Acad. Sci. USA *99*, 5919–5924.

Wustmann, G., Szidonya, J., Taubert, H., and Reuter, G. (1989). The genetics of position-effect variegation modifying loci in *Drosophila melanogaster*. Mol. Gen. Genet. *217*, 520–527.

Ye, Q., Callebaut, I., Pezhman, A., Courvalin, J.C., and Worman, H.J. (1997). Domain-specific interactions of human HP1-type chromodomain proteins and inner nuclear membrane protein LBR. J. Biol. Chem. *272*, 14983–14989.

Zhang, Z., Hayashi, M.K., Merkel, O., Stillman, B., and Xu, R.M. (2002). Structure and function of the BAH-containing domain of Orc1p in epigenetic silencing. EMBO J. *21*, 4600–4611.

Accession Numbers

The atomic coordinates of the HP1β CSD-EMSY-N complex have been deposited in the RCSB Protein Data Bank with the accession code 2FMM.