BIOCHEMISTRY
including biophysical chemistry & molecular biology

ARTICLE

pubs.acs.org/biochemistry

Identification of Nucleic Acid Binding Residues in the FCS Domain of the Polycomb Group Protein Polyhomeotic

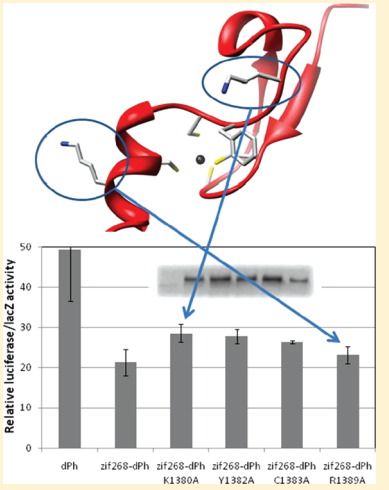
Renjing Wang,[†] Udayar Ilangovan,[†] Belinda Z. Leal,[†] Angela K. Robinson,[†] Barbara T. Amann,[‡] Corey V. Tong,[†] Jeremy M. Berg,[§] Andrew P. Hinck,[†] and Chongwoo A. Kim*[,†]

[†]Department of Biochemistry, University of Texas Health Science Center at San Antonio, MSC 7760, 7703 Floyd Curl Drive, San Antonio, Texas 78229-3990, United States

[‡]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 9/1N101 Memorial Drive, MSC 0921, Bethesda, Maryland 20892-0921, United States

[§]National Institute of General Medical Sciences, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, United States

ABSTRACT: Polycomb group (PcG) proteins maintain the silent state of developmentally important genes. Recent evidence indicates that noncoding RNAs also play an important role in targeting PcG proteins to chromatin and PcG-mediated chromatin organization, although the molecular basis for how PcG and RNA function in concert remains unclear. The Phe-Cys-Ser (FCS) domain, named for three consecutive residues conserved in this domain, is a 30−40-residue Zn^{2+} binding motif found in a number of PcG proteins. The FCS domain has been shown to bind RNA in a non-sequence specific manner, but how it does so is not known. Here, we present the three-dimensional structure of the FCS domain from human Polyhomeotic homologue 1 (HPH1, also known as PHC1) determined using multidimensional nuclear magnetic resonance methods. Chemical shift perturbations upon addition of RNA and DNA resulted in the identification of Lys 816 as a potentially important residue required for nucleic acid binding. The role played by this residue in Polyhomeotic function was demonstrated in a transcription assay conducted in *Drosophila* S2 cells. Mutation of the Arg residue to Ala in the *Drosophila* Polyhomeotic (Ph) protein, which is equivalent to Lys 816 in HPH1, was unable to repress transcription of a reporter gene to the level of wild-type Ph. These results suggest that direct interaction between the Ph FCS domain and nucleic acids is required for Ph-mediated repression.



The Polycomb group (PcG) is a family of gene silencing proteins that repress important developmental regulator genes, including homeotic (HOX) genes. Once the expression of these genes is no longer required, the PcG proteins maintain the silent state over many cell divisions. In stem cells, members of the PcG repress genes that promote differentiation, thus playing an important role in maintaining the pluripotency of these cells.[1,2]

The PcG functions at the level of chromatin, although the precise mechanism by which PcG complexes bind to chromatin has not been fully elucidated. While the presence of MBT and chromo domains within PcG proteins can allow direct association with methylated histones, DNA not bound to histones is also important for PcG function. For example, the *Drosophila* PcG protein Pleiohomeotic (Pho), the only PcG protein that houses a specific DNA binding domain, acts in concert with the multiprotein PcG complex called Polycomb Repression Complex 1 (PRC1) to associate with specific gene regulatory sequences called Polycomb response elements (PREs).[3,4] Pho is also a component of a different PcG complex called PhoRC that includes the PcG protein dSfmbt.[5] A single complex containing both a specific DNA binding protein (Pho) and the ability to bind methylated histone through the MBT domains of dSfmbt[6] may utilize a combinatorial approach to bind specific chromatin sites that have both the Pho binding element and methylated histones.

RNA is also emerging as an important player in PcG-mediated repression. PRC2, a multiprotein PcG complex that catalyzes the trimethylation of histone H3 K27, associates with a noncoding RNA (ncRNA) called Xist, facilitating the inactivation of the X chromosome in female mammals.[7] This interaction occurs between a region within Xist consisting of 7.5 repeats of a 28-nucleotide sequence (RepA) and the PRC2 component Ezh2, resulting in the proper targeting of PRC2 to the inactive X (Xi) chromosome.[8] RING1B, a PRC1 component, is targeted to the Xi chromosome in a manner that is dependent on Xist but independent of PRC2.[9] Recent reports have revealed that many other ncRNAs can associate with PRC2. Long ncRNAs HOTAIR and Kcnq1oT_1 are associated with PRC2 and, like Xist, play an important role in targeting the methyl transferase activity of PRC2 to specific chromatin locations.[10,11] Moreover, a global

Received: September 14, 2010
Revised: January 19, 2011
Published: February 25, 2011

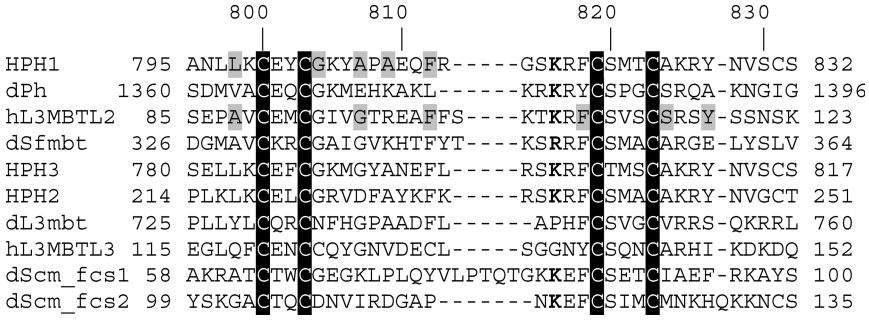


Figure 1. ClustalW[42] alignment of FCS sequences. The metal binding Cys residues are highlighted in black. The lightly shaded amino acids are buried residues (>75% buried surface area) in the two FCS domain structures that have been determined. The prefixes d and h in the protein names indicate *Drosophila melanogaster* and human proteins, respectively. The two tandem FCS domains in the *Drosophila* Scm protein are indicated by fcs1 and fcs2.

analysis of ncRNA interactions revealed that approximately 20% of all mammalian ncRNAs are associated with PRC2.[12]

RNAi-related proteins have also been shown to be involved in PcG-mediated repression.[13] The *Drosophila* RNAi proteins Dicer-2, Piwi, and Argonaute were found to colocalize with PRC1 components Polycomb (Pc) and Polyhomeotic (Ph) and were required for the maintenance of long-range chromosomal interactions. Interestingly, small transcripts from an exogenous PRE were detected. These small RNAs appear to be required for maintaining the long-range interactions because mutant flies unable to make the transcripts could not mediate these interactions.

A domain found in a number of PcG proteins that is a candidate for directly binding RNA is the Phe-Cys-Ser (FCS) domain. In vitro, the FCS domain can bind both RNA and DNA in a manner independent of the sequence of the nucleic acids.[14] The focus of that study was a *Caenorhabditis elegans* PcG protein called SOP-2. The RNA binding ability of SOP-2 allows its localization into large nuclear bodies often termed "PcG bodies".[13,15] PcG bodies are thought to correspond to sites where clustering of PREs and gene silencing occur through long-range interactions between the PREs. SOP-2 does not contain an FCS domain but rather utilizes a different sequence of amino acids for RNA association. The conserved RNA binding function of the FCS domain and the RNA binding region of SOP-2 was demonstrated by a SOP-2 chimera that replaced its RNA binding regions with the mouse Ph homologue FCS domain that was able to partially restore the ability to localize into PcG bodies.[14]

The FCS domain is a 30−40-residue sequence that is able to chelate a Zn^{2+} atom via four conserved Cys residues (Figure 1). A number of PcG proteins contain an FCS domain, including all Ph orthologs, *Drosophila* Sex comb on midleg (Scm) where two FCS domains are present as a tandem repeat, *Drosophila* Sfmbt, and *Drosophila* lethal 3 malignant brain tumor l3mbt along with two of its related mammalian proteins, l3mbt-like 2 (L3MBTL2) and L3MBTL3. A recent structure of the FCS domain from human L3MBTL2 determined using multidimensional NMR methods revealed an architecture that is compatible with RNA binding.[16] In this structure, three mostly conserved, positively charged residues are on one face of the structure and were suggested as potential functional residues that could bind the negatively charged backbone of RNA. As an alternative to a nucleic acid binding role, a recent study has implicated the Scm and dSfmbt FCS domains as being involved in protein−protein interactions.[6]

To gain insight into the molecular basis of Ph, and also PcG function, we have biochemically characterized and determined the solution structure of the FCS domain from human Ph homologue 1 (HPH1, also known as PHC1). Our results strongly suggest that hindering the ability of the Ph FCS domain to bind nucleic acids also hinders the ability of Ph to repress transcription.

EXPERIMENTAL PROCEDURES

Protein Preparation. The HPH1 FCS domain (residues 783−828) was cloned into a modified pET-3c (Novagen) plasmid containing an N-terminal hexahistidine tag followed by a TEV protease cleavage site. The HPH1 FCS domain was expressed in BL21-Gold (DE3) cells (Stratagene) that had been pretransformed with the pRARE plasmid (Novagen). Bacterial cells from a 1 L culture were resuspended in 10 mL of 50 mM Tris (pH 8.0), 100 mM NaCl, 25 mM imidazole (pH 7.5), 1 mM PMSF, 10 mM β-mercaptoethanol, and 5% glycerol. Cells were lysed by sonication and purified using Ni^{2+} affinity chromatography. The eluted protein was digested with TEV to remove the N-terminal sequence followed by a second Ni^{2+} affinity chromatography run in which the nonbinding fractions were collected. The protein was further purified by ion exchange chromatography. The final polypeptide after TEV cleavage contained an N-terminal GTR sequence followed by the HPH1 residues.

UV−Visible Spectroscopy. The bacterially expressed and purified HPH1 FCS was stripped of metal by lowering the pH through the addition of trifluoroacetic acid and then purified by reversed-phase HPLC. In an anerobic environment, the lyophilized, metal-free HPH1 FCS domain was resuspended in 100 mM Hepes and 50 mM NaCl (pH 7.0) to a final concentration of 66.7 μM. One equivalent of $CoCl_2$ was added followed by measurement of the visible spectrum. One equivalent of $ZnCl_2$ was then added to the solution for the observation of any bleaching of the spectrum.

NMR Spectroscopy: Structure Determination, Dynamics, and Chemical Shift Perturbations. Stable isotope-labeled protein samples (containing either ^{15}N alone or both ^{15}N and ^{13}C) were expressed in minimal medium containing isotopically labeled substrates, purified as described above, and prepared at a concentration of 1.5 mM in 10 mM $NaPO_4$ (pH 6.0) and 50 mM NaCl. Fractionally labeled (10% ^{13}C) samples were prepared in minimal medium containing a ratio of ^{13}C-labeled to unlabeled glucose of of 0.3 g/L:2.7 g/L. Two-dimensional $^{13}C-^{1}H$ HSQC spectra of biosynthetic fractionally ^{13}C- and ^{1}H-labeled protein were used for stereospecific assignment of the side chain methyl groups (valine and leucine)[17,18] and the aromatic (Phe and Tyr) groups.[19] NMR experiments were performed at 300 K using

Table 1. Structural Restraints and Statistics for 20 Lowest-Energy Structures of the HPH1 FCS Domain

Structural Restraints	
total no. of NOE distance restraints	1024
no. of NOE distance restraints	
intraresidue ($\|i-j\|=0$)	374
sequential ($\|i-j\|=1$)	285
short-range ($2<\|i-j\|<5$)	146
long-range ($\|i-j\|>5$)	220
no. of dihedral restraints (extracted from TALOS[24])	
φ	26
ψ	26
no. of RDC restraints	
D_{NH}	28
D_{CH}	28
no. of coupling restraints	
$^3J_{HNH\alpha}$	32
no. of hydrogen bond restraints	9
no. of zinc coordinate restrains	6
Structural Statistics	
root-mean-square deviation (rmsd) from ideal geometry (±SD)	
bond lengths (Å)	0.0046 ± 0.00024
bond angles (deg)	0.8473 ± 0.039
improper angels (deg)	2.125 ± 0.019
average atomic rmsd from mean structure (±SD)	
residues 795−813 and 815−828 (backbone)	0.46
residues 795−813 and 815−828 (heavy atoms)	1.05
Structure Evaluation (Ramachandran plot) (residues 783−828)	
residues in most favored regions (%)	62.4
residues in additional allowed regions (%)	30.4
residues in generally allowed regions (%)	5.2
residue in disallowed regions (Asp793) (%)	2.0

Bruker 600 and 700 MHz spectrometers fitted with either conventional (700 MHz) or cryogenically cooled (600 MHz) 5 mm ^{1}H probes equipped with ^{13}C and ^{15}N decoupler and pulsed field gradient coils. All spectra were processed and analyzed with NMRPipe[20] and NMRView.[21] Structure calculations were performed with CNS version 1.1[22] using ARIA version 1.2,[23] incorporating NOEs and dihedral angle restraints calculated with TALOS.[24] The structure was refined using the $^3J_{HNH\alpha}$ couplings, RDCs, hydrogen bonds, and the tetrahedral Zn^{2+} geometry restraints (Table 1). Backbone amide HSQC-based longitudinal (T_1) and transverse (T_2) relaxation times and heteronuclear Overhauser effects ($\{^1H\}-^{15}N$ NOE) were recorded using standard Bruker pulse programs. Time scale estimations of the HPH1 FCS internal conformational dynamics were determined by analyzing the ^{15}N relaxation parameters using the Lipari−Szabo model-free formalism[25,26] in ModelFree version 4.0 (http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/software/modelfree.html). The chemical shift assignments have been deposited in the BMRB (accession number 17396) and the Protein Data Bank (PDB) (entry rcsb102085). The coordinates have been deposited as PDB entry 2L8E.

For the measurement of the chemical shift perturbations to the HPH1 FCS domain in the presence of nucleic acids, the following oligonucleotides were used: 34-nucleotide RNA (5′-GGGC-ACGCGUAUUGCCCUAGUGGCCGGCGUGCCC-3′) and 14 bp dsDNA (5′-CAGCCATATGGCTG-3′) with its reverse complement.

All oligonucleotides were placed into the same buffer as the HPH1 FCS domain prior to the titration experiment. The HPH1 FCS peptide was titrated with increasing amounts of nucleic acids. A $\{^1H\}-^{15}N$ HSQC spectrum was recorded after each titration point. The weighted average chemical shift changes of the assigned residues were calculated using the equation $\Delta_{av} = [(\Delta\delta_{NH}^2 + \Delta\delta_N^2/25)/2]^{1/2}$.

The dissociation constant (K_d) was estimated by globally fitting the shifts for different residues to a common K_d value using the following equation as described by Lian and Roberts:[27]

$$\Delta\delta_{obs} = \frac{\delta_b - \delta_f}{2P_T}\left[P_T + L_T + K_d - \sqrt{(P_T + L_T + K_d)^2 - 4P_T L_T}\right]$$

$\Delta\delta_{obs}$ is the observed difference in weighted average chemical shifts between the free state and that observed at each titration point, $\delta_b - \delta_f$ is the total weighted average chemical shift difference between the bound and free states, and P_T and L_T are the total concentrations of protein and ligand (RNA), respectively. Pro Fit software (QuantumSoft) was used for the nonlinear least-squares fitting.

Transcription Assay. On day 1 of the assay, the following plasmids were transfected into 1×10^5 *Drosophila* S2 cells using the Fugene HD transfection reagent (Roche Applied Science): (1) 100 ng of the *Drosophila Ph* expression plasmid under control of a constitutive actin 5c promoter (kind gift from A. J. Courey), (2) 7.5 ng of a *lacZ* gene expression plasmid, also under control of the actin 5c promoter (kind gift from Y. Shiio), and (3) 7.5 ng of pGL2-Basic with three tandem repeats of the zif268 DNA binding sites cloned immediately upstream of a metallothionine promoter (MTp) that controls expression of the *luciferase* gene. On day 3, the expression of the *luciferase* gene was induced via addition of $CuSO_4$ (100 μM). On day 4, cells were harvested and lysed using 100 mM potassium phosphate buffer (pH 7.8), 0.2% Triton X-100, and 0.5 mM DTT. For all individual transfections, equal volumes of lysate were used for the Dual-Light Combine Reporter Gene Assay System (Applied Biosystems) to measure both luciferase and β-galactosidase activities. The data are presented as the ratio of the two enzyme activities.

RESULTS

Tetrahedral Coordination of Metal by the HPH1 FCS Domain. The ability of the FCS domain to bind Zn^{2+} was originally suggested on the basis of the presence of four conserved Cys residues whose sulfide groups were predicted to be the chelating residues for the zinc metal. If so, the four thiolates would be arranged in a tetrahedral geometry around the Zn^{2+}. To utilize the tetrahedral geometric constraints in our structure calculations, we needed to first confirm that the FCS domain does indeed bind Zn^{2+} with tetrahedral geometry. We measured the UV−visible spectrum of the HPH1 FCS domain (residues 783−828 of HPH1) fully substituted with Co^{2+}. The d−d electron transitions of the Co^{2+} that is ligated through four ligands arranged tetrahedrally produce a distinct spectrum, while the Zn^{2+}-bound FCS domain is spectroscopically silent. The solution of the Co^{2+}-bound HPH1 FCS domain displays a

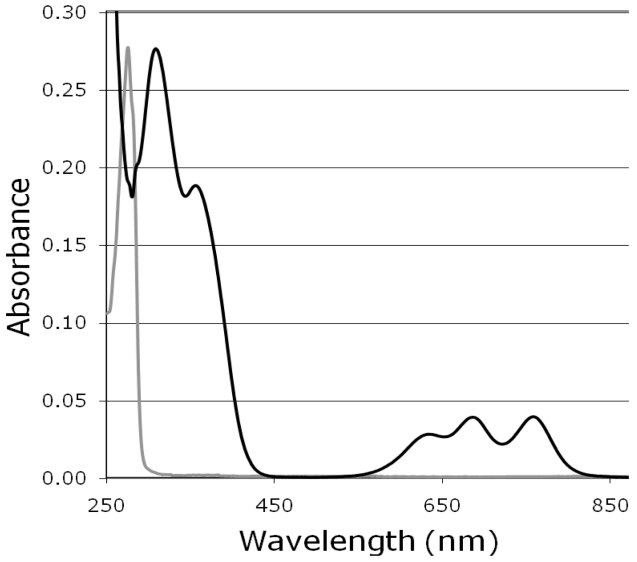


Figure 2. UV−visible spectrum of the HPH1 FCS domain bound to Co^{2+}. The black spectrum is the spectrum with Co^{2+} bound and the gray spectrum that after the addition of $ZnCl_2$.

clear blue color with three distinct d−d transition peaks at 635, 686, and 759 nm in the visible spectrum (Figure 2). In addition, the Cys-S to Co^{2+} charge transfer bands are observed between 275 and 450 nm. This spectrum resembles the UV−visible spectrum of other Cys4-coordinated Zn^{2+} binding peptides when bound to Co^{2+},[28] including a treble-clef zinc finger peptide that is structurally similar to the L3MBTL2 FCS domain.[29] Our results are consistent with a structure in which the HPH1 FCS domain does indeed bind Zn^{2+} with tetrahedral geometry through four cysteine bonds. Upon addition of Zn^{2+}, all cobalt transitions were bleached as a consequence of substitution of Zn^{2+} for Co^{2+}.

Determination of the NMR Structure of the HPH1 FCS Domain. Our initial $\{^1H\}-^{15}N$ HSQC measurement of the HPH1 FCS domain showed a spectrum with a well-dispersed pattern of backbone amide signals whose total number was nearly equal to the anticipated value (Figure 3A). We, thus, proceeded to prepare a ^{13}C- and ^{15}N-labeled HPH1 FCS sample to assign the backbone and side chain resonances and to determine the three-dimensional structure. The structural restraints used and statistics are summarized in Table 1. The ensemble of the 10 calculated lowest-energy structures consistent with the NMR structural restraints is shown in Figure 3B. HPH1 residues 783−795 of our construct appear to be disordered as this region lacks sufficient structural restraints required to define a consistent structure. The ordered region of the HPH1 FCS structure [Asn 797−Asn 828 (Figure 3C)] consists of an antiparallel β-sheet (Asn 797−Pro 808), followed by a loop region (Ala 809−Ser 820) and then an α-helix (Met 821−Asn 828). Zn^{2+} binding residues Cys 800 and Cys 803 are in the β-sheet region; Cys 819 is in the loop, and Cys 823 is in the α-helix. As with other similar zinc binding peptides,[29] zinc binding appears to be critical for maintaining the fold of the FCS domain as Zn^{2+} chelation is responsible for determining the orientation of the HPH1 FCS α-helix with respect to the β-sheet. The structure of the HPH1 FCS domain closely resembles that of the FCS domain of L3MBTL2[16] (rmsd of 1.7 Å^2 over 28 Cα atoms) and the *Escherichia coli* YacG zinc binding domain[30] (rmsd of 2.3 Å^2 over 29 Cα atoms), with the greatest variation in structures occurring in the loop region of the structures (Figure 3D).

HPH1 FCS Dynamics Indicates an Ordered Loop Region. Aside from the disordered N-terminal region of our construct, the central loop region (Ala 809−Ser 820) within the FCS domain exhibited increased rmsds among the ensemble of low-energy structures as compared to the regions of defined secondary structure (Figure 3B). As the restraint density is lower in this region than in the regions of regular secondary structure (18 restraints per residue as compared to 27 in the β-sheet and α-helix), it is unclear whether the increased rmsd is due to this or intrinsic flexibility of the loop. We investigated this in greater detail by measuring the T_1, T_2, and $\{^1H\}-^{15}N$ NOE (HNNOE) relaxation parameters (Figure 4A). The long T_1 and T_2 times along with the low HNNOE values for HPH1 FCS N-terminal residues 783−797 indicate significant flexibility in this region on the picosecond to nanosecond time scale, consistent with disorder of this region. We conducted a more detailed analysis of the molecular motions of the FCS domain using the Lipari−Szabo extended model-free formalism[25,26] to distinguish motions on different time scales (Figure 4B). HPH1 residues 783−797, as expected, have low S^2 and high τ_e values, indicating that this region is much more flexible on the picosecond to nanosecond time scale than the rest of the structure. The loop, residues 809−820, appears to be rigid on the picosecond to nanosecond time scale as it exhibits comparably high S^2 values. However, several residues in the loop, including Glu 810, Gly 814, Lys 816, and Arg 817, exhibit elevated R_{ex} values indicating greater mobility of these residues on the microsecond to millisecond time scale. Interestingly, the residue with the highest R_{ex} and, thus, greatest level of motion on the microsecond to millisecond time scale is Lys 816, a residue that was identified as playing an important role in nucleic acid recognition (see below). We conclude from these results that the loop is rigid on the picosecond to nanosecond time scale but flexible on the microsecond to millisecond time scale.

Identification of Nucleic Acid Binding Residues of the HPH1 FCS Domain. The Ph FCS domain can bind nucleic acids in a non-sequence specific manner.[14] In the HPH1 FCS domain used in this study, there are several positively charged residues within the ordered HPH1 FCS domain as well as two in the disordered N-terminus. These residues could contact the negatively charged phosphate backbone of nucleic acids. Moreover, there are three mostly conserved basic residues (Lys 816, Arg 817, and Lys 825) that are clustered together on both the HPH1 FCS and L3MBTL2 FCS structures,[16] suggesting a potential conserved role for these residues. If indeed the function of the FCS domain is to bind nucleic acids, albeit in a manner independent of the nucleic acid sequence, then it can be predicted that a particular set of FCS domain residues, and not a random collection of positively charged residues, would be more inclined to perform this function.

To identify the residues that contact nucleic acids, we measured the perturbations of the HPH1 FCS backbone amide resonances upon addition of RNA and DNA. Because the FCS domain can bind RNA and DNA with no sequence specificity, we arbitrarily chose a 34-nucleotide RNA predicted to form a stem−bulge−loop structure (Figure 5A, inset) and 14 bp double-stranded DNA. A $\{^1H\}-^{15}N$ HSQC spectrum was recorded at several titration points for each of the nucleic acid titration experiments (Figure 5). The HPH1 FCS domain titration with the 34-nucleotide RNA showed a number of residues whose

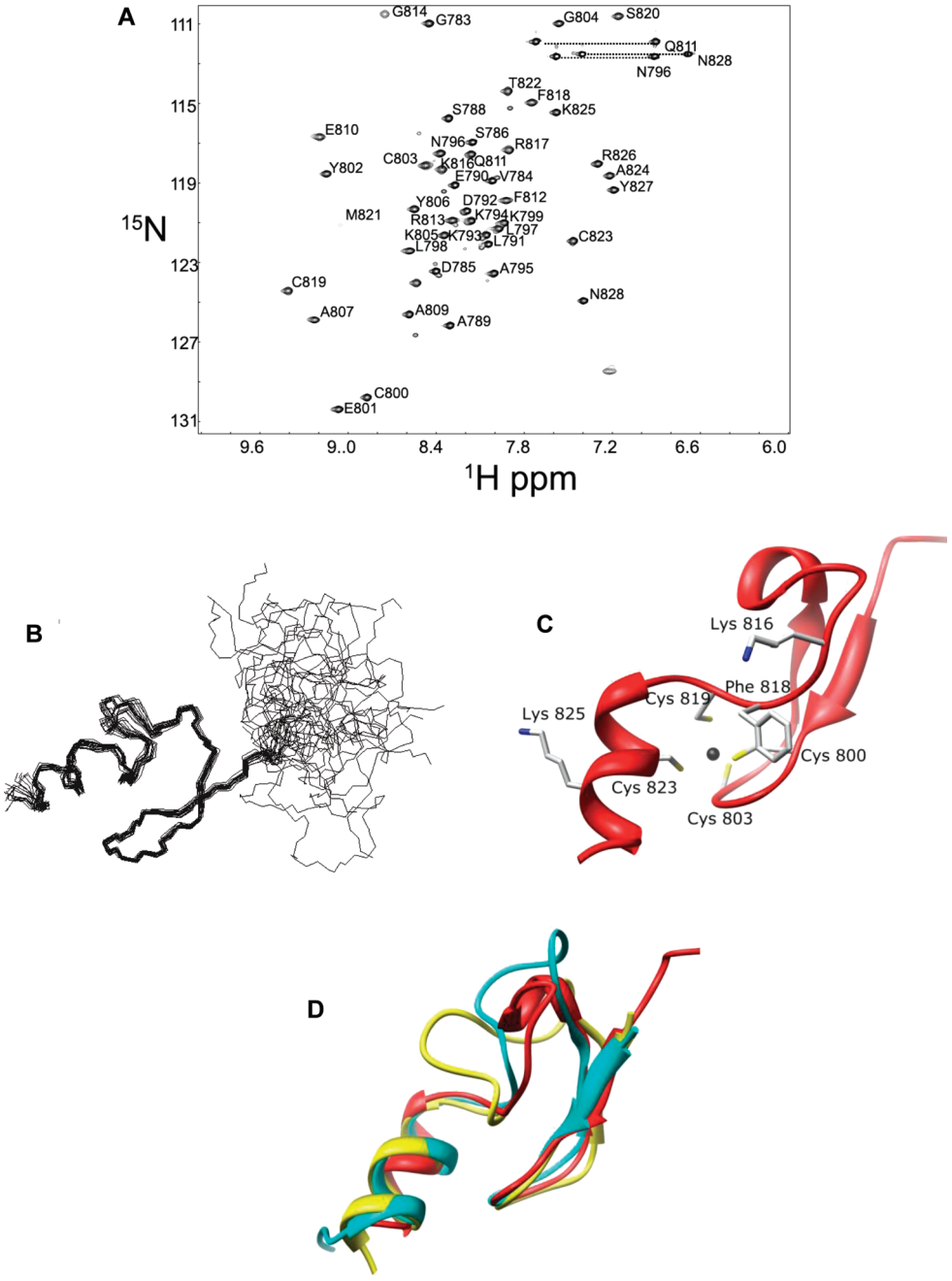


Figure 3. Determination of the solution structure of the HPH1 FCS domain. (A) HPH1 FCS $\{^1H\}-^{15}N$ HSQC. The chemical shift assignment for each residue is indicated. (B) Ensemble of the 20 lowest-energy structures of the HPH1 FCS domain. The rmsd of the secondary structure backbone is 0.46. (C) Ribbon diagram of the HPH1 FCS domain. The four metal binding residues are highlighted along with other residues that were mutated for the transcription assay. (D) Overlay of HPH1 FCS (red), L3MBTL2 FCS (yellow, PDB entry 2W0T),[16] and the bacterial YacG peptide (cyan, PDB entry 1LV3).[30] The overlay shows the greatest variability of this fold occurs in the loop region.

HSQC signals were altered but not appreciably broadened and thus fell under the fast exchange regime (Figure 5A). Most notable were Lys 816, Phe 818, and either Cys 803 or Ser 815 (the identity of this backbone amide signal could not be discerned because these two residues are degenerate) (Figure 5B). Titration with the 14 bp double-stranded DNA was less remarkable than titration with the 34-nucleotide RNA (Figure 5C). However, all the residues whose signals were altered in the presence of the DNA are also perturbed with RNA (Glu 810, Phe 812, Arg 813, Lys 816, and Phe 818). All these residues are in the loop region of the HPH1 FCS domain. The results of these titration experiments are analyzed by the backbone N−H maximum weighted average shift chemical differences (Figure 5D). A charged surface representation of the FCS structure (Figure 5E) clearly shows the proximity of the loop residues whose chemical shifts show the greatest perturbation. Lys 825, which was previously described as being part of a potential nucleic acid binding surface on the L3MBTL2 FCS structure[16] but does not appear to be involved in binding, is quite distant from the loop. From these experiments, we conclude the FCS domain does indeed use a specific set of

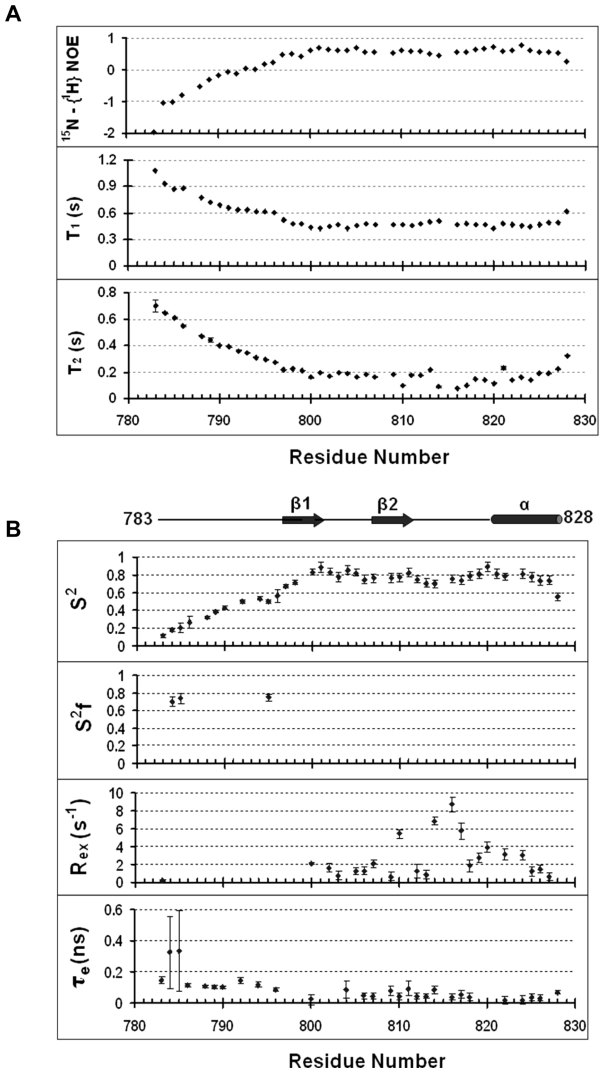


Figure 4. Backbone conformational dynamics of the HPH1 FCS domain. (A) Longitudinal (T_1), transverse (T_2), and $\{^1H\}-^{15}N$ NOE values. The errors in individual T_1 and T_2 measurements were estimated by Monte Carlo simulations. (B) Lipari−Szabo model-free results calculated from ^{15}N T_1, ^{15}N T_2, and $\{^1H\}-^{15}N$ NOE data. The modeling was performed at a τ of 2.85 ns. The calculated Lipari−Szabo S^2, S_f^2, τ_e, and R_{ex} parameters are shown. Residues that could not be fit to a specific motional model were not assigned S_f^2, τ_e, or R_{ex} values.

residues to contact nucleic acids and that the loop region appears to be especially important for binding.

The affinity between the HPH1 FCS domain and the 34-nucleotide RNA was calculated using the equation described by Lian and Roberts[27] (see Experimental Procedures for the equation). This equation relates the observed chemical shift difference ($\Delta\delta_{obs}$) to the dissociation constant (K_d) and the total concentrations of the protein and RNA (ligand), P_T and L_T, respectively. The maximum weighted average chemical shift changes of four residues with the greatest changes in their chemical shifts (Glu 810, Ser 815, Lys 816, and Phe 818) were plotted as a function of the total concentrations of protein and RNA concentrations, resulting in an estimated K_d of 3.0 ± 1.1 mM (Figure 5F). The changes in chemical shifts were insufficient to estimate a similar K_d with the 14 bp DNA. This likely indicates a weaker affinity for the 14 bp DNA because of insufficient amounts of the intact FCS−DNA complex that were allowed for detection of perturbed chemical shifts.

Mutation of the Nucleic Acid Binding Residues Hinders Ph Transcription Repression Ability. We next determined the consequence on gene silencing of disturbing the potential interaction of the Ph FCS domain with nucleic acids. For this purpose, we utilized a transcription assay conducted in *Drosophila* S2 cells using the full-length *Drosophila* Ph (dPh) protein. There is abundant evidence showing *Drosophila* PcG proteins function in a manner very similar to that of their mammalian counterparts. Most compellingly, mammalian PcG proteins have been used to rescue *Drosophila* phenotypes resulting from mutant orthologs,[31−33] including dPh.[34] In our transcription assay, we fused the DNA binding domain of human zif268 to the N-terminus of a variety of dPh proteins to target the zif268 fused chimeric proteins to three zif268 binding sites immediately upstream of the metallothionine promoter (MTp) that controls expression of the *luciferase* gene. On the basis of the results of our in vitro binding studies with the HPH1 FCS domain, we introduced four different mutations into the dPh FCS domain and determined the ability of these mutant dPh proteins, compared to that of wild-type dPh, to repress expression of the *luciferase* gene. The four individual mutations we introduced into dPh were Lys1380Ala (equivalent to HPH1 Lys 816), Tyr1382Ala (equivalent to HPH1 Phe 818), Cys1383Ala (equivalent to HPH1 Cys 819), and Arg1389Ala (equivalent to HPH1 Lys 825). Cys 1383 is a metal binding residue and would be expected to disrupt the structure of the FCS domain. Arg 1389 is equivalent to HPH1 Lys 825 and is mostly conserved as a positively charged residue in the sequence alignment (Figure 1). However, unlike the other conserved positively charged residue, HPH1 Lys 816, Lys 825 displays minimal HSQC signal perturbation upon addition of RNA and DNA (Figure 5). This residue was, therefore, chosen as a control for the mutation experiments, anticipating that if nucleic acid binding is required for Ph-mediated repression, then a Lys825Ala mutant would not be defective in its repressive abilities.

The results of the transcription assay are shown in Figure 6. zif268-fused wild-type dPh and mutant dPh proteins exhibit substantial repression activity compared to dPh not fused to zif268 and thus not targeted to the MTp, indicating that dPh can repress transcription with a nonfunctional FCS domain. However, some of the mutants exhibited a slight but significant reduction in their repressive abilities. Lys1380Ala, Tyr1382Ala, and Cys1383Ala, mutations expected to disrupt the structure or nucleic acid binding ability of the FCS domain, exhibited less ability to repress transcription compared to the levels of the wild type. Intriguingly, dPh Cys1383Ala, which we expected to fully destabilize the FCS domain, was slightly better at repressing transcription than dPh Lys1380Ala, though its ability is within the standard deviation. One possibility for this outcome is that the mutation does not fully unfold the FCS structure. To determine whether this was the case, we introduced the equivalent mutation into the HPH1 FCS domain, Cys819Ala, and measured its one-dimensional ^1H NMR spectrum. To our surprise, HPH1 FCS Cys819Ala does appear to possess some tertiary structural features as evidenced by the presence of both up- and downfield chemical shifts (Figure 6B). Additionally, the stability of the FCS domain may increase further inside the nucleus upon binding of nucleic acids. The Arg1383Ala mutant, whose HPH1 Lys 825 counterpart appears to play a lesser role in nucleic acid binding (Figure 5), showed no difference in

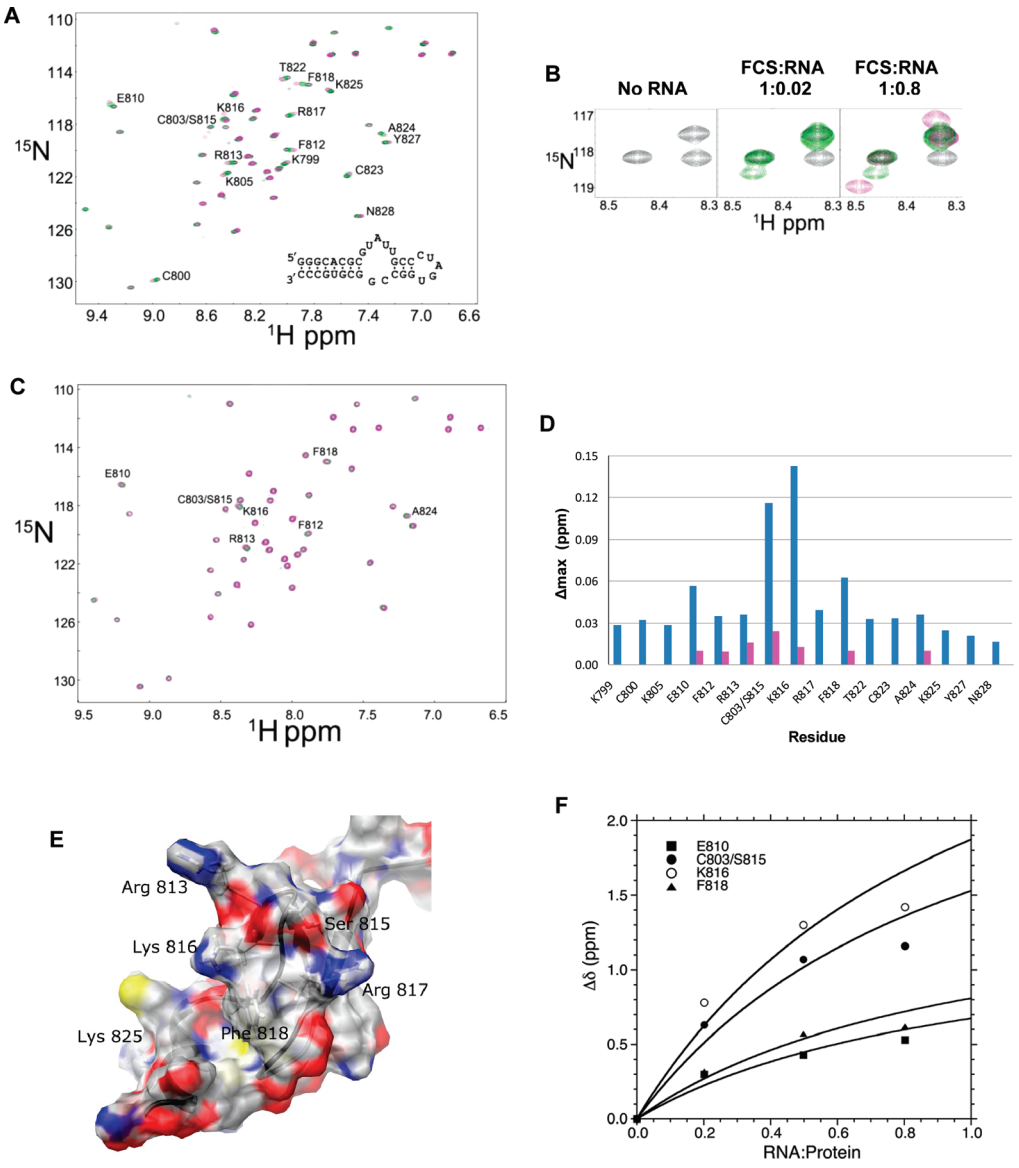


Figure 5. Nucleic acid titration of the HPH1 FCS domain. (A) Overlay of the HSQC spectra for HPH1 FCS titrated with the 34-nucleotide RNA. The black, green, and purple HSQC spectra correspond to spectra after the addition of 0, 0.2, and 0.8 molar equiv, respectively, of the 34-nucleotide RNA to the ^{15}N-labeled HPH1 FCS domain. The inset shows the predicted secondary structure of the 34-nucleotide RNA used in the titration. (B) Expanded view of the section of the HSQC spectrum that includes the backbone amide signals for Cys 803/Ser 815 and Lys 816. These two signals show the largest perturbation in the presence of the 34-nucleotide RNA. Unfortunately, the signals for Cys 803 and Ser 815 could not be discerned. (C) Overlay of the spectra for HPH1 FCS titrated with 0, 0.5, and 2 molar equiv (black, green, and purple, respectively) of the 14 bp dsDNA. Residues that show perturbation in the presence of the DNA are highlighted. (D) Backbone N−H weighted average chemical shift differences of the HPH1 FCS domain upon binding the 34-nucleotide RNA (blue) and the 14 bp dsDNA (purple). The maximum weighted average chemical shift difference, Δ_{max}, was calculated from chemical shift values obtained from the $\{^1H\}-^{15}N$ HSQC spectra recorded in the absence of nucleic acids and the highest, saturated, concentration. (E) Charged surface representation of the HPH1 FCS domain. The structure is oriented like Figure 3C. Several of the loop residues as well as Lys 825 are highlighted. (F) HPH1 FCS domain−34-nucleotide RNA binding curves of the four residues that show the largest chemical shift perturbation in the titration experiment. The data were fit to the equation described by Lian and Roberts[27] (see Experimental Procedures).

repression ability compared to the wild type. Although Lys 825 is not directly binding nucleic acid, it may simply contribute by providing a more positive charge environment for nucleic acid binding. These results, which nicely correlate our in vitro nucleic acid binding studies with the transcription repression assay conducted in S2 cells, further confirm the important role played by Lys 1380 (or HPH1 Lys 816) in nucleic acid binding that is required for full repression mediated by dPh.

DISCUSSION

These are the key findings of our experiments: (1) determination of the three-dimensional structure of the HPH1 FCS domain using multidimensional NMR methods, (2) identification of the FCS domain loop region, in particular HPH1 Lys 816, as being important for binding RNA, and (3) demonstrating that the in vitro nucleic acid binding ability of the Ph FCS domain correlates with the ability of Ph to repress transcription.

dx.doi.org/10.1021/bi101487s |*Biochemistry* 2011, 50, 4998–5007

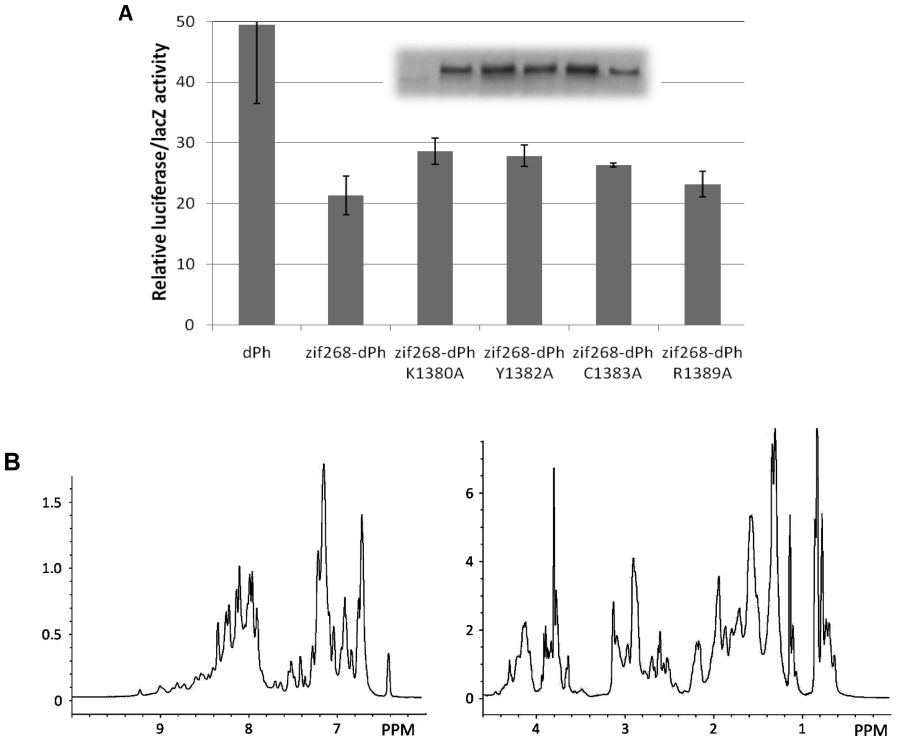


Figure 6. Transcription repression assay. (A) Error bars show the standard deviation of the results from three independent transfections. The inset shows an immunoblot against the Flag epitope-tagged Ph proteins demonstrating relatively equal amounts of the zif368-fused Ph proteins expressed in the S2 cells. This experiment was repeated several times, and in every case, the results were consistent with what is shown. (B) One-dimensional ^{1}H NMR spectrum of the HPH1 FCS Cys819Ala mutant.

Two different NMR experiments indicate that the loop region of the FCS domain is important for nucleic acid binding. The perturbations of the HPH1 FCS HSQC signal in the presence of RNA are greatest for the loop residues. HPH1 Lys 816, along with the conserved Phe 818, shows substantial movement of its resonances in the presence of the 34-nucleotide RNA. The signals of other nearby loop residues (Phe 812, Arg 813, and Arg 816) also exhibited perturbations in the presence of nucleic acids (Figure 5). In addition to the movement of the backbone amide signals in the HSQC spectra, it is also interesting that our backbone conformational dynamics study indicated that residues in the loop exhibit higher mobility on the microsecond to millisecond time scale compared to the β-sheet or α-helix regions of the structure (Figure 4B). Such flexibility within the loop of the FCS domain could be utilized to bind to specific targets via conformational selection.[35] If so, the flexible loop region preexisting in a number of different conformations may allow the FCS domain to bind to different target nucleic acids with high affinity. A recent study has revealed that the more flexible regions within the ubiquitin structure are correlated with the regions that are involved in protein−protein interactions due to conformational selection.[36] Therefore, the greater flexibility of the loop region is likely important for recognition of the nucleic acid by the FCS domain. The loop region is also the most variable section of the FCS domain, in both sequence (Figure 1) and structure, as it is the section within the FCS domains from HPH1 and L3MBTL2 where the difference is the greatest in the overlay of the two structures (Figure 3D). Because the loop appears to be important for nucleic acid recognition, it may be possible that the different loop structures among the FCS domains allow recognition of different nucleic acid sequences and/or structures, thereby providing additional specificity determinants.

An additional factor that could influence the specificity and affinity for the binding reactions involving the Ph FCS domain is its oligomeric state. All Ph orthologs contain a C-terminal sterile α motif (SAM) domain, while dSfmbt, Scm, and the l3mbt proteins contain MBT domains. In vitro, the *Drosophila* Ph SAM domain can self-associate into a helical polymer architecture;[37] the dSfmbt and Scm MBT domains may also form higher-order stoichiometries.[6,38,39] Ph oligomerization via the SAM domain would result in the presence of multiple FCS domains that would potentially contribute to the specificity and affinity of the nucleic acids that bind to the FCS domain. The presence of the FCS, MBT, and SAM domains in the same PcG protein is common. *Drosophila* Scm contains not only two adjacent FCS domains but also a polymeric SAM domain and an MBT domain. *Drosophila* Sfmbt and l3mbt as well as mammalian L3MBTL3 have both FCS and SAM domains as well as an MBT domain. It will be of interest to determine whether oligomerization related to the presence of the SAM or MBT domains influences the function of the FCS domain.

If multimerization of the FCS domain occurs via the protein−protein interactions with the other parts of the protein, we might expect that the affinity between a single isolated FCS domain would be relatively weak. Consistent with this, the chemical shifts in the HSQC spectrum do not broaden and remain in fast exchange during the titration with nucleic acids. In addition, the K_d of the HPH1 FCS domain−34-nucleotide RNA fusion was estimated to be 3 mM, and even weaker with DNA. It is important to point out that there is currently no RNA molecule or DNA sequence that is known to be a specific binding partner

for any of FCS domain. While the affinity for a specific binding partner would obviously increase, we speculate that the ability of the FCS domain to bind weakly and nonspecifically to many different sequences of DNA and structures of RNA may be how it precisely evolved to function to organize different nucleic acid elements around a higher oligomeric display of FCS domains.

The results presented here show that Ph proteins that contain mutations that hinder FCS domain−nucleic acid interactions are less able to repress transcription. The molecular role of the FCS domain in Ph-mediated repression remains to be determined. However, in view of the recently recognized participation of ncRNA in PcG-mediated represssion, it is possible that the FCS domain binds ncRNAs, which in turn would provide a targeting function, allowing Ph and other FCS domain-containing proteins to associate at specific chromatin locations. Alternatively, the FCS domain may function to bind DNA directly. If so, the FCS domain would need to bind to DNA that is free of histones. PREs are susceptible to histone replacement and have a reduced level of histones in chromatin immunoprecipitation experiments.[40,41] Furthermore, PRC1 can form a complex with PRE DNA in a manner that would only be possible in the absence of histones.[4] These results point to the ability of PRC1 to form a complex with naked DNA. Binding of the FCS domain to DNA may help facilitate the formation of such complexes. Future studies of the FCS domain will help provide insights into these possibilities.

AUTHOR INFORMATION

Corresponding Author

*E-mail: chong@biochem.uthscsa.edu. Phone: (210) 567-8779. Fax: (210) 567-8778.

Funding Sources

This work was supported by the American Heart Association (0830111N), the American Cancer Society (RSG-08-285-01-GMC), and the Department of Defense Breast Cancer Research Program (BC075278) (C.A.K.). Support for the NMR facility was provided by the University of Texas Health Science Center at San Antonio (UTHSCSA) and National Institutes of Health Grant NCI P30 CA54174 (CTRC at UTHSCSA).

ACKNOWLEDGMENT

We thank Drs. P. John Hart and Susan T. Weintraub for comments about the manuscript. We thank Dr. Jens Wöhnert for providing the RNA used in this study.

ABBREVIATIONS

FCS domain, phenylalanine-cysteine-serine domain; PcG, Polycomb group; PRC1 and PRC2, Polycomb repression complexes 1 and 2, respectively; PRE, Polycomb response element; Ezh2, enhancer of zeste homologue 2; Ph, Polyhomeotic; HPH1, human Polyhomeotic homologue 1; Scm, sex comb on midleg; L3MBTL2, lethal 3 malignant brain tumor-like 2; Sfmbt, Scm-like with four MBT domains; NMR, nuclear magnetic resonance; HSQC, heteronuclear single-quantum coherence; HNNOE, $\{^1H\}-^{15}N$ NOE; RDC, residual dipolar coupling; HOX, Homeotic; ncRNA, noncoding RNA; Xi, inactivated X chromosome; SAM, sterile α motif; TEV, tobacco etch virus; MBT domain, malignant brain tumor domain.

REFERENCES

(1) Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A., and Jaenisch, R. (2006) Polycomb Complexes Repress Developmental Regulators in Murine Embryonic Stem Cells. *Nature 441*, 349–353.

(2) Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K., Koseki, H., Fuchikami, T., Abe, K., Murray, H. L., Zucker, J. P., Yuan, B., Bell, G. W., Herbolsheimer, E., Hannett, N. M., Sun, K., Odom, D. T., Otte, A. P., Volkert, T. L., Bartel, D. P., Melton, D. A., Gifford, D. K., Jaenisch, R., and Young, R. A. (2006) Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. *Cell 125*, 301–313.

(3) Mohd-Sarip, A., Cleard, F., Mishra, R. K., Karch, F., and Verrijzer, C. P. (2005) Synergistic Recognition of an Epigenetic DNA Element by Pleiohomeotic and a Polycomb Core Complex. *Genes Dev. 19*, 1755–1760.

(4) Mohd-Sarip, A., van der Knaap, J. A., Wyman, C., Kanaar, R., Schedl, P., and Verrijzer, C. P. (2006) Architecture of a Polycomb Nucleoprotein Complex. *Mol. Cell 24*, 91–100.

(5) Klymenko, T., Papp, B., Fischle, W., Kocher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Muller, J. (2006) A Polycomb Group Protein Complex with Sequence-Specific DNA-Binding and Selective Methyl-Lysine-Binding Activities. *Genes Dev. 20*, 1110–1122.

(6) Grimm, C., Matos, R., Ly-Hartig, N., Steuerwald, U., Lindner, D., Rybin, V., Muller, J., and Muller, C. W. (2009) Molecular Recognition of Histone Lysine Methylation by the Polycomb Group Repressor dSfmbt. *EMBO J. 28*, 1965–1977.

(7) Plath, K., Fang, J., Mlynarczyk-Evans, S. K., Cao, R., Worringer, K. A., Wang, H., de la Cruz, C. C., Otte, A. P., Panning, B., and Zhang, Y. (2003) Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science 300*, 131–135.

(8) Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J., and Lee, J. T. (2008) Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. *Science 322*, 750–756.

(9) Schoeftner, S., Sengupta, A. K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., and Wutz, A. (2006) Recruitment of PRC1 Function at the Initiation of X Inactivation Independent of PRC2 and Silencing. *EMBO J. 25*, 3110–3122.

(10) Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., and Chang, H. Y. (2007) Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell 129*, 1311–1323.

(11) Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008) Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. *Mol. Cell 32*, 232–246.

(12) Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B. E., van Oudenaarden, A., Regev, A., Lander, E. S., and Rinn, J. L. (2009) Many Human Large Intergenic Noncoding RNAs Associate with Chromatin-Modifying Complexes and Affect Gene Expression. *Proc. Natl. Acad. Sci. U.S.A. 106*, 11667–11672.

(13) Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U., and Cavalli, G. (2006) RNAi Components are Required for Nuclear Clustering of Polycomb Group Response Elements. *Cell 124*, 957–971.

(14) Zhang, H., Christoforou, A., Aravind, L., Emmons, S. W., van den Heuvel, S., and Haber, D. A. (2004) The *C. elegans* Polycomb Gene SOP-2 Encodes an RNA Binding Protein. *Mol. Cell 14*, 841–847.

(15) Saurin, A. J., Shiels, C., Williamson, J., Satijn, D. P., Otte, A. P., Sheer, D., and Freemont, P. S. (1998) The Human Polycomb Group Complex Associates with Pericentromeric Heterochromatin to Form a Novel Nuclear Domain. *J. Cell Biol. 142*, 887–898.

(16) Lechtenberg, B. C., Allen, M. D., Rutherford, T. J., Freund, S. M., and Bycroft, M. (2009) Solution Structure of the FCS Zinc Finger

Domain of the Human Polycomb Group Protein L(3)Mbt-Like 2. *Protein Sci. 18*, 657–661.

(17) Neri, D., Szyperski, T., Otting, G., Senn, H., and Wuthrich, K. (1989) Stereospecific Nuclear Magnetic Resonance Assignments of the Methyl Groups of Valine and Leucine in the DNA-Binding Domain of the 434 Repressor by Biosynthetically Directed Fractional ^{13}C Labeling. *Biochemistry 28*, 7510–7516.

(18) Senn, H., Werner, B., Messerle, B. A., Weber, C., Traber, R., and Wüthrich, K. (1989) Stereospecific Assignment of the Methyl ^{1}H NMR Lines of Valine and Leucine in Polypeptides by Nonrandom ^{13}C Labelling. *FEBS Lett. 249*, 113–118.

(19) Jacob, J., Louis, J. M., Nesheiwat, I., and Torchia, D. A. (2002) Biosynthetically Directed Fractional ^{13}C Labeling Facilitates Identification of Phe and Tyr Aromatic Signals in Proteins. *J. Biomol. NMR 24*, 231–235.

(20) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR 6*, 277–293.

(21) Johnson, B. A., and Blevins, R. A. (1994) NMR View: A Computer Program for the Visualization and Analysis of NMR Data. *J. Biomol. NMR 4*, 603–614.

(22) 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., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr. D54* (Part 5), 905–921.

(23) Linge, J. P., O'Donoghue, S. I., and Nilges, M. (2001) Automated Assignment of Ambiguous Nuclear Overhauser Effects with ARIA. *Methods Enzymol. 339*, 71–90.

(24) Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein Backbone Angle Restraints from Searching a Database for Chemical Shift and Sequence Homology. *J. Biomol. NMR 13*, 289–302.

(25) Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc. 104*, 4546–4559.

(26) Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *J. Am. Chem. Soc. 104*, 4559–4570.

(27) Lian, L. Y., and Roberts, G. C. K. (1993) in *NMR of Macromolecules: A Practical Approach* (Roberts, G. C. K., Ed.) pp 153, Oxford University Press, New York.

(28) Ghering, A. B., Shokes, J. E., Scott, R. A., Omichinski, J. G., and Godwin, H. A. (2004) Spectroscopic Determination of the Thermodynamics of Cobalt and Zinc Binding to GATA Proteins. *Biochemistry 43*, 8346–8355.

(29) Seneque, O., Bonnet, E., Joumas, F. L., and Latour, J. M. (2009) Cooperative Metal Binding and Helical Folding in Model Peptides of Treble-Clef Zinc Fingers. *Chemistry 15*, 4798–4810.

(30) Ramelot, T. A., Cort, J. R., Yee, A. A., Semesi, A., Edwards, A. M., Arrowsmith, C. H., and Kennedy, M. A. (2002) NMR Structure of the *Escherichia coli* Protein YacG: A Novel Sequence Motif in the Zinc-Finger Family of Proteins. *Proteins 49*, 289–293.

(31) Muller, J., Gaunt, S., and Lawrence, P. A. (1995) Function of the Polycomb Protein is Conserved in Mice and Flies. *Development 121*, 2847–2852.

(32) Gorfinkiel, N., Fanti, L., Melgar, T., Garcia, E., Pimpinelli, S., Guerrero, I., and Vidal, M. (2004) The *Drosophila* Polycomb Group Gene Sex Combs Extra Encodes the Ortholog of Mammalian Ring1 Proteins. *Mech. Dev. 121*, 449–462.

(33) Atchison, L., Ghias, A., Wilkinson, F., Bonini, N., and Atchison, M. L. (2003) Transcription Factor YY1 Functions as a PcG Protein in Vivo. *EMBO J. 22*, 1347–1358.

(34) Ringrose, L., and Paro, R. (2004) Epigenetic Regulation of Cellular Memory by the Polycomb and Trithorax Group Proteins. *Annu. Rev. Genet. 38*, 413–443.

(35) Boehr, D. D., Nussinov, R., and Wright, P. E. (2009) The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol. 5*, 789–796.

(36) Lange, O. F., Lakomek, N. A., Fares, C., Schroder, G. F., Walter, K. F., Becker, S., Meiler, J., Grubmuller, H., Griesinger, C., and de Groot, B. L. (2008) Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science 320*, 1471–1475.

(37) Kim, C. A., Gingery, M., Pilpa, R. M., and Bowie, J. U. (2002) The SAM Domain of Polyhomeotic Forms a Helical Polymer. *Nat. Struct. Biol. 9*, 453–457.

(38) Nagem, R. A., Dauter, Z., and Polikarpov, I. (2001) Protein Crystal Structure Solution by Fast Incorporation of Negatively and Positively Charged Anomalous Scatterers. *Acta Crystallogr. D57*, 996–1002.

(39) Gil, J., Bernard, D., Martinez, D., and Beach, D. (2004) Polycomb CBX7 has a Unifying Role in Cellular Lifespan. *Nat. Cell Biol. 6*, 67–72.

(40) Papp, B., and Muller, J. (2006) Histone Trimethylation and the Maintenance of Transcriptional ON and OFF States by trxG and PcG Proteins. *Genes Dev. 20*, 2041–2054.

(41) Mito, Y., Henikoff, J. G., and Henikoff, S. (2007) Histone Replacement Marks the Boundaries of Cis-Regulatory Domains. *Science 315*, 1408–1411.

(42) Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003) Multiple Sequence Alignment with the Clustal Series of Programs. *Nucleic Acids Res. 31*, 3497–3500.