

# Identification of Methyllysine Peptides Binding to Chromobox Protein Homolog 6 Chromodomain in the Human Proteome\*<sup>§</sup>

Nan Li<sup>‡§</sup>, Richard S. L. Stein<sup>‡§</sup>, Wei He<sup>‡§</sup>, Elizabeth Komives<sup>‡</sup>, and Wei Wang<sup>‡¶||</sup>

**Methylation is one of the important post-translational modifications that play critical roles in regulating protein functions. Proteomic identification of this post-translational modification and understanding how it affects protein activity remain great challenges. We tackled this problem from the aspect of methylation mediating protein–protein interaction. Using the chromodomain of human chromobox protein homolog 6 as a model system, we developed a systematic approach that integrates structure modeling, bioinformatics analysis, and peptide microarray experiments to identify lysine residues that are methylated and recognized by the chromodomain in the human proteome. Given the important role of chromobox protein homolog 6 as a reader of histone modifications, it was interesting to find that the majority of its interacting partners identified via this approach function in chromatin remodeling and transcriptional regulation. Our study not only illustrates a novel angle for identifying methyllysines on a proteome-wide scale and elucidating their potential roles in regulating protein function, but also suggests possible strategies for engineering the chromodomain–peptide interface to enhance the recognition of and manipulate the signal transduction mediated by such interactions. *Molecular & Cellular Proteomics* 12: 10.1074/mcp.O112.025015, 2750–2760, 2013.**

Post-translational modification (PTM)<sup>1</sup> is essential for the proper biological function of many proteins in almost every organism. Biochemistry and proteomics studies have identified a series of PTMs (1), including phosphorylation (2), glycosylation (3), acetylation (4, 5), ubiquitinylation (6), sumoylation (7, 8), and methylation (9–11). These PTMs relate to a wide range of important biological processes and functions,

such as transcriptional regulation, signal transduction (12), tumorigenesis, and protein degradation. A great variety of PTMs on histones (*i.e.* histone modifications) can affect the expression of target genes by directly altering chromatin structure or recruiting regulatory protein complexes (13–15). A series of histone modification readers have been identified that recognize histone modifications and transmit the signal to the next layer in the protein interaction network (16, 17). In addition to the discovery of those histone modification readers, structural biologists have solved several crystal structures of histone modification readers and demonstrated the characteristics of various binding pockets that recognize different histone modifications (16).

Polycomb protein comprises a family of histone modification readers that specifically recognize methylated H3K9/H3K27 (18, 19). It was first identified in *Drosophila melanogaster* as a heterochromatin-associated protein and a component of polycomb repressive complex 1 (PRC1), which is related to gene repression (20). A series of proteins also identified as polycomb group proteins are the components of PRC1 and PRC2 and involve chromatin-related gene repression. Further studies have shown that polycomb group proteins, including polycomb itself, are involved in a wide range of gene regulation activities related to development, sex-related regulation, tumorigenesis, cell cycle control, and cell differentiation (21–25). However, little is known about the function of polycomb beyond its role as a polycomb group protein.

Five human homologs of *Drosophila* polycomb have been found: CBX2, CBX4, CBX6, CBX7, and CBX8 (18, 26). Complex structures between CBX proteins and methylated histone peptides H3K9me3, H3K27me2, and H3K27me3 (27–30) have been resolved. All of those complex structures show the same binding site structure as that of polycomb, in which a three-residue aromatic cage composed of tryptophan and either tyrosine or phenylalanine is located on the interface, and the side chain of methylated lysine extends into the aromatic cage. The cation- $\pi$  interaction between the three aromatic rings and the positively charged methylated lysine becomes the main stabilizing force for the complex.

Recent studies have also shown that non-histone lysine methylation exists widely in the human proteome. Levy *et al.* discovered 118 proteins that can be methylated by lysine

From the <sup>‡</sup>Department of Chemistry and Biochemistry, 9500 Gilman Drive, University of California, San Diego, La Jolla, California 92093-0359; <sup>¶</sup>Department of Cellular and Molecular Medicine, 9500 Gilman Drive, University of California, San Diego, La Jolla, California 92093-0359

Received October 19, 2012, and in revised form, July 5, 2013

Published, MCP Papers in Press, July 10, 2013, DOI 10.1074/mcp.O112.025015

<sup>1</sup> The abbreviations used are: CBX6, chromobox protein homolog 6; PRC, polycomb repressive complex; PTM, post-translational modification; SASA, solvent-accessible surface area.

methyltransferase SETD6 using fluorescent methods and 114 others via radioactive methods; 26 proteins were detected with both methods (31). Dhayalan *et al.* described 91 non-histone peptide substrates that can be methylated by lysine methyltransferase SET7/9; among them, 9 proteins were further verified *in vivo* (32). Additionally, Balasubramanian *et al.* showed that SET7/9 can also methylate FXR (33). The known function for lysine methylation on non-histone proteins is very limited. Only a few non-histone methylated proteins have been well studied. One example is tumor repressor p53, which has multiple lysine methylation sites, such as K370 (34), K372 (35, 36), K373 (37), and K382 (38, 39). p53K372me stabilizes p53 protein in the nucleus, which is required for the transcription activation of downstream genes (35, 36), whereas p53K370me and p53K382me repress the p53 regulation effect on target genes (34, 39). A second example is the lysine methylation of DNA methyltransferase DNMT1 (40, 41); DNMT1-K142me can lead to its degradation (42). Furthermore, RELA, a component of the pleiotropic transcription factor NF- $\kappa$ B, can be monomethylated by SET6D at K310, resulting in a repressive effect for the RELA-driven processes (43). Despite these discoveries, the proteomic identification of methylation by means of mass spectrometry remains nontrivial because of the lack of effective enrichment methods for specific methylated peptides. In addition, determining which proteins recognize methyllysine is an important but very challenging problem that has not been widely addressed for non-histone proteins.

Several computational methods related to domain-peptide interaction are available, but only on limited domains, such as SH2 (44), SH3 (45–49), and PDZ (50–55). For all of these domains, enough interaction data have already been obtained to train a reasonably good machine learning model. However, there is no effective way to obtain machine learning models for interactions without sufficient experimental data. There are some prediction methods for potential lysine methylation sites (56–59), but the prediction results give information only on whether lysine sites can be methylated; the relationships among methylated lysine sites, proteins, and their functions remain unclear.

In this study, we first characterized the domain-peptide interface using a mutation matrix generated from molecular dynamics and virtual mutagenesis, which revealed residues critical for binding and suggested possible ways to engineer the recognition interface. We then searched for methyllysine-containing peptides in the human proteome that may bind to the chromobox protein homolog 6 (CBX6) chromodomain. We exploited a computational pipeline to integrate diverse data, including the binding affinity estimated by the mutation matrix, sequence conservation and secondary structure of the peptide, subcellular location of the protein containing the candidate peptide, mass spectrometry signal of methylation, and co-expression between CBX6 and its putative binder. The peptides selected by this computational pipeline were next

examined via peptide microarray binding assay, and 50 non-histone peptides were confirmed as CBX6 binders. Gene ontology analysis of these CBX6 interacting partners showed that the majority of them function in chromatin modification and transcriptional regulation. Our study provides a novel way to systematically identify methylated peptides in the proteome, illustrate the protein-protein interactions mediated by such methylation, and thus significantly expand our limited knowledge of non-histone protein methylations.

#### MATERIALS AND METHODS

**CBX6 Modeling Procedure**—The crystal structure for the CBX6-H3K27me3 complex was obtained from the Protein Data Bank (60) (PDB ID 3I90). The models were prepared for molecular dynamics simulation with the AMBER 9.0 (61) module LEAP. The trimethylated lysine, which was not recognized by the standard residue library of LEAP, was treated as a nonstandard residue. Its atomic charges were calculated with the HF/6–31G\* basis set of Gaussian03 (62) and fitted subsequently with RESP (63). [Supplemental Table S1](#) lists the resulting partial charges on the trimethylated lysine atoms. In the Gaussian03 charge calculation, the trimethylated lysine residue was capped (64) by ACE and NME residues so that it could be treated as a complete molecule. The FF03 force field (65) was used to handle all atoms in the model except for the trimethylated zeta nitrogen of H3K27, for which the GAFF force field (66) was used. The CBX6-histone model was solvated with a rectangular, periodic box of TIP3P water molecules (67) that extended 10 Å past the solute on all sides. Chloride counter ions were added to neutralize the models.

**Molecular Dynamics Simulation**—The CBX6-H3K27me3 system was minimized in two stages. In the first stage, the water molecules and counter ions were relaxed while organic molecules were restrained by a harmonic force of 100 kcal/mol Å<sup>2</sup>. In the second stage, the restraints were removed, and the entire system was allowed to minimize. Each stage consisted of 400 steps of steepest descent followed by 1600 steps of conjugate gradient minimization. After the minimization stages, the system was heated from 0 to 300 K over 30 ps. The temperature of 300 K and a pressure of 1 atm were maintained by a Berendsen thermostat and barostat (68) with a coupling time of 0.2 ps. The SHAKE algorithm (69) was used to restrain all bonds involving hydrogen. Initial temperatures were assigned randomly at the beginning of the heating run according to the Maxwell-Boltzmann distribution. The system was equilibrated for 200 ps and then underwent a production run lasting 4 ns. The CBX6 heavy atoms were restrained by a harmonic force of 10 kcal/mol Å<sup>2</sup> in order to ensure the stability of the complex during the equilibration and production runs.

**Virtual Mutagenesis**—We formed the mutation matrix by identifying the 21 most critical binding residues in CBX6 for binding to H3K27me3, according to Table I. Each of these residues was mutated to all of the other 19 naturally occurring amino acids. This produced  $21 \times 19 = 399$  separate models, each a point mutation of the wild-type complex between CBX6 and histone H3. In a separate operation, the sequence of the protein was preserved and each of the H3 residues featured in the model, except for H3K27, was mutated to each of the other 19 amino acids. This produced another  $10 \times 19 = 190$  separate models. The software program SCAP (70) was used to perform the point mutations. The standard procedure of minimization, heating, and equilibration was performed on each of the models, in addition to 2 ns of production run. Binding free energies were calculated from the second half (the 1.0–2.0 ns interval) of the production run for each model.

**Trajectory Analysis, Free Energy Calculation, and Decomposition Analysis**—The root mean square deviation for each system during the equilibration and production run phases of the molecular dynamics simulation was calculated by the PTRAJ module in the AMBER package (61). As a representative sample, the root mean square deviation for the CBX6/H3K27me3 wild type is shown in Fig. 1. Based on the stabilization of the root mean square deviation after about 1.0 ns, the interval for the binding free energy calculation was chosen for the latter half of each production run, between 1.0 and 2.0 ns. MM-GBSA (71–73) was used to calculate the binding free energies for each CBX6-histone complex using the MM-PBSA module in AMBER 9.0, according to the general procedure given by Wang and Kollman (74) and Hou *et al.* (73). The binding free energies  $\Delta G_{\text{bind}}$  were calculated from 100 snapshots taken at equal intervals between the 1.0-ns and 2.0-ns marks of the production run trajectory.  $\Delta G_{\text{bind}}$  was calculated as follows:

$$\Delta G_{\text{bind}} = \langle G_{\text{complex}} \rangle - (\langle G_{\text{protein}} \rangle + \langle G_{\text{peptide}} \rangle) \quad (\text{Eq. 1})$$

Here,  $\langle G_{\text{complex}} \rangle$ ,  $\langle G_{\text{protein}} \rangle$ , and  $\langle G_{\text{peptide}} \rangle$  are the respective individual free energies of the complex, protein, and peptide, and each term is calculated by summing the contributions from the electrostatic potential  $E_{\text{elec}}$ , van der Waals potential  $E_{\text{vdw}}$ , and solvation free energy  $G_{\text{solv}}$ . In turn,  $G_{\text{solv}}$  is the sum of the polar contribution  $G_{\text{polar}}$  and nonpolar contribution  $G_{\text{nonpolar}}$  to the solvation free energy.

$$G_{(\text{complex/protein/peptide})} = E_{\text{elec}} + E_{\text{vdw}} + G_{\text{solv}} \quad (\text{Eq. 2})$$

$$G_{\text{solv}} = G_{\text{polar}} + G_{\text{nonpolar}} \quad (\text{Eq. 3})$$

$G_{\text{polar}}$  was calculated by setting the value of IGB to 2, activating the generalized Born parameters of Onufriev *et al.* (75).  $G_{\text{nonpolar}}$  was estimated as 0.0072 times the solvent-accessible surface area (SASA), as measured via the LCPO method (76). A grid size of 0.5 Å was used to solve the Poisson–Boltzmann equation, and the probe radius (77) was set at 1.4 Å. A value of 4.0 was used for the interior dielectric constant, which applied to the region of the protein–histone binding pocket, and the external dielectric constant was given throughout as 80.0.

To investigate the molecular basis of recognition specificity, we conducted component decomposition of the binding free energy. The contributions to the binding free energy were also broken down for individual residues and by component (van der Waals, electrostatic, and solvation) for each of the complexes. Decompositions were performed by invoking the DECOMP module in the MM-GBSA program within AMBER. Each H3 or CBX6 residue was evaluated according to its overall contribution to the binding interaction, as opposed to a pairwise analysis of every CBX6/H3 residue combination.

**Mutation Score**—Every candidate peptide can be considered as a mutant from the H3K27me3 peptide. Therefore, a site-specific mutation matrix was used to calculate the mutation score, which is the estimated difference in binding energy between the candidate peptide and H3K27me3.

$$\text{Mutation score} = \sum_{\substack{i=1 \\ i \neq 9}}^{11} M(i, A_i) \quad (\text{Eq. 4})$$

The number represented by  $i$  is the amino acid position on the peptide;  $i$  ranges from 1 to 11, and position 9 must be a lysine that is assumed to be methylated.  $M$  is the mutation matrix generated from Virtual Mutagenesis analysis, and  $A_i$  is the  $i$ th residue of the candidate peptide.

**Conservation Score**—To evaluate the conservation of each candidate peptide, we first retrieved 42 vertebrate proteomes (supplemental Table S2) from Ensemble (78). For each human protein containing

a candidate peptide, the homologous proteins in the other proteomes were found using NCBI BLAST (79). Multiple sequence alignment was then generated for the 43 homologous proteins using CLUSTALW2 (80). The overall conservation score was calculated from the multiple sequence alignment using the BLOSUM62 matrix (81).

$$\text{Overall conservation score} = \sum_{\substack{i=1 \\ i \neq 9}}^{11} \sum_{m=1}^{42} \text{BLOSUM}(A_i, A_i^m) \quad (\text{Eq. 5})$$

In Equation 5,  $A_i$  is the  $i$ th amino acid in the candidate peptide and  $A_i^m$  is the  $i$ th amino acid in the homolog of  $m$ th vertebrate proteome. In addition to the conservation of the entire peptide, the lysine residue to be methylated should also be highly conserved because of the functional constraint, which can be evaluated using the multiple sequence alignment as well. The cutoff values for the overall conservation and lysine conservation scores were chosen as 2200 and 37, respectively.

**Mass Spectrometry Evidence**—About 30 million mass spectra from human tissues were downloaded from PRIDE (82) and PeptideAtlas (83). The complete collection of human proteins in the Uniprot/Swiss-prot database was used as the reference database for the MS scan. The scanning programs XITANDEM (84) and INSPECT (85) were used to perform the scan, with parent mass error set to 1 Da and fragment mass error to 0.2 Da. We searched for peptides containing mono-, di-, or trimethylated lysine. The scanned results from the two programs were then merged.

**Secondary Structure**—The H3K27 peptide adopts an extended conformation ( $\beta$  strand and loop) in the CBX6–H3K27 complex, and it is clear that the peptide cannot fit to the binding pocket in helical form. Therefore, we removed the candidate peptides that form a helix. For proteins with available structures in the Protein Data Bank, DSSP (86) was used to define the secondary structure for each residue. For proteins without solved structures, two secondary structure prediction methods, PSI-PRED3.0 (87) and SSpro4.0 (88), were used to estimate the secondary structure for each residue. If a residue was predicted as a helix by either method, its secondary structure was specified as helical. Candidate peptides containing more than two helix residues between the 4th and the 10th position were removed.

**Solvent-accessible Surface Area**—The SASA was calculated for all structure-available peptides. If no structure was available, to avoid noise, the SASA filter was not applied. The difference in SASA for H3K27 peptide before and after binding was calculated using DSSP (86). The SASA in every candidate peptide for each residue was calculated using DSSP and compared with that of H3K27. We required that each residue from the 2nd to the 10th position have more than 70% of the SASA of the corresponding H3K27 residue.

**Co-expression between CBX6 and Its Putative Interacting Partners**—Data for human microarray experiments were downloaded from the Stanford Microarray Database (89). The ratio between red and green channel intensities was used as the gene expression level. If a gene was represented by multiple probes, the average expression value was used. First, the expression value of each gene was converted to a Z-score such that these values were comparable across different microarray experiments. When calculating the pairwise correlation of two genes using the 10,882 microarray experiments, we discarded experiments in which either gene's expression was missing. Then, mutual information was calculated to evaluate the co-expression of the two genes. The Unigene ID was taken from the Stanford Microarray Database directly as the identifier for each gene. The mapping of the Unigene ID to the Uniprot accession ID was performed by parsing the Uniprot XML file. In cases when the Uniprot accession ID corresponded to multiple Unigene IDs, all of the IDs

were considered as one gene, and their expression values were averaged before normalization. The correlation coefficients of expression levels between CBX6 and all available genes from the Stanford Microarray Database were calculated and used to construct a background distribution, based on which a *p* value was computed for a given co-expression coefficient. We took a loose *p* value cutoff of 0.5 as the co-expression filter threshold.

**Peptide and Protein Synthesis and Screening**—The chromodomain of CBX6 protein (amino acids 11–69) was expressed as a GST-tagged fusion protein and purified as described in Ref. 90. The purity of the protein was examined via SDS-PAGE electrophoresis followed by both Coomassie staining and Western blot using an anti-GST-HRP conjugate (Santa Cruz Biotechnology, Santa Cruz, CA). The concentration of purified protein was determined via BCA assay (Amresco). A total of 248 methylated peptides were synthesized by Sigma Aldrich (desalted, mass-spectrometry checked). The peptides were then printed as triplets onto glass slides (ArrayIt), together with a Cy3 marker and an anti-GST (mouse monoclonal) antibody (Thermo) as references.

The peptide microarray was rinsed with TBST buffer (25 mM Tris, 125 mM NaCl, 0.05% Tween-20, pH 8) and then underwent blocking with 5% nonfat milk in TBST (room temperature for 1 h or 4 °C overnight). The slide was then incubated at 4 °C with CBX6-GST fusion protein at a final concentration of 5 μM in 5% nonfat milk/TBST for 12 h. After being washed three times for 10 min each time with TBST, the slide was incubated with an anti-GST mouse monoclonal IgG antibody (Thermo) at a final concentration of 1 μg/ml in 5% nonfat milk/TBST and shaken gently for 1 h at room temperature. A secondary anti-mouse IgG Dylight-488 conjugated antibody (Thermo) was added to a final concentration of 0.1 μg/ml after three more cycles of 10-min washing with TBST. The slide was shaken for 1 h at room temperature and washed three times with TBST.

**Data Acquisition and Analysis in Peptide Microarray Experiments**—The dried microarray slide was scanned using a Hamamatsu Nano-Zoomer 2.0HT Slide Scanning System (Neuroscience Light Microscopy Facility, University of California, San Diego). Data quantification was processed using the microarray processing software ImageJ (91), for which the fluorescence intensity of a microarray spot was defined as its own mean intensity minus the background mean intensity around it on the scanned image.

**Pull-down of CBX6 by Methylated Peptides**—Methylated peptides (Sigma Aldrich) (0.05 to 0.1 mg) were incubated in PBS (10 mM, pH 7.2) with dry N-hydroxysuccinimide-activated resin (Thermo) to allow covalent coupling of the peptide via the C' terminal amine. The resin was then washed abundantly three times with PBS (pH 7.2), and the excessive reactive sites were quenched with 1 M Tris for 1 h. The resin was washed three times with TBS (25 mM, pH 8) and incubated with CBX6 chromodomain overnight at 4 °C. The resin was then washed thoroughly six times with TBS, boiled with Laemmli buffer (DTT), and analyzed with SDS-PAGE gel. The CBX6 chromodomain was detected via Western blotting using an anti-GST-HRP conjugate (Santa Cruz Biotechnology, Santa Cruz, CA).

RESULTS

**Deciphering the Recognition of Methylated Peptides by the CBX6 Chromodomain**—In order to reveal the atomic mechanisms of CBX6-peptide recognition, we performed residue decomposition (see “Materials and Methods”) on the CBX6–H3K27me3 complex to identify the residues critical for binding (Table I). The two residues making the greatest contribution to the binding interaction are Phe11 (–3.78 kcal/mol) and Trp32 (–4.77 kcal/mol). These residues are two of the three in the aromatic, hydrophobic cage that interacts with H3K27me3

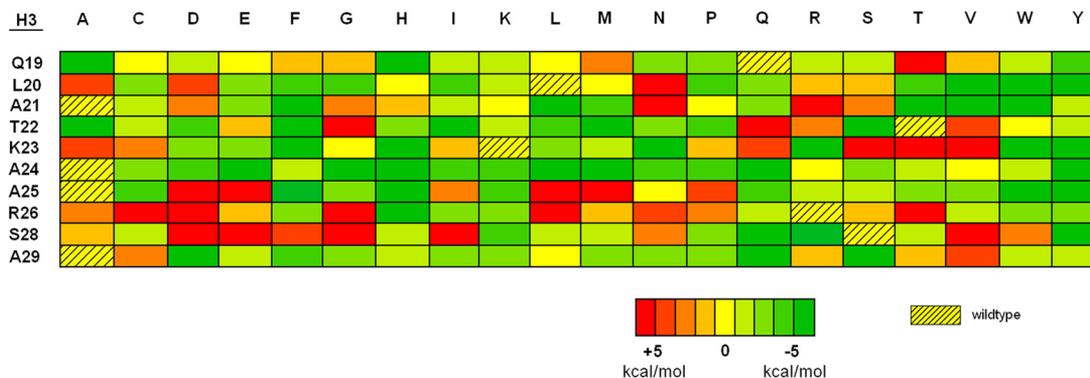
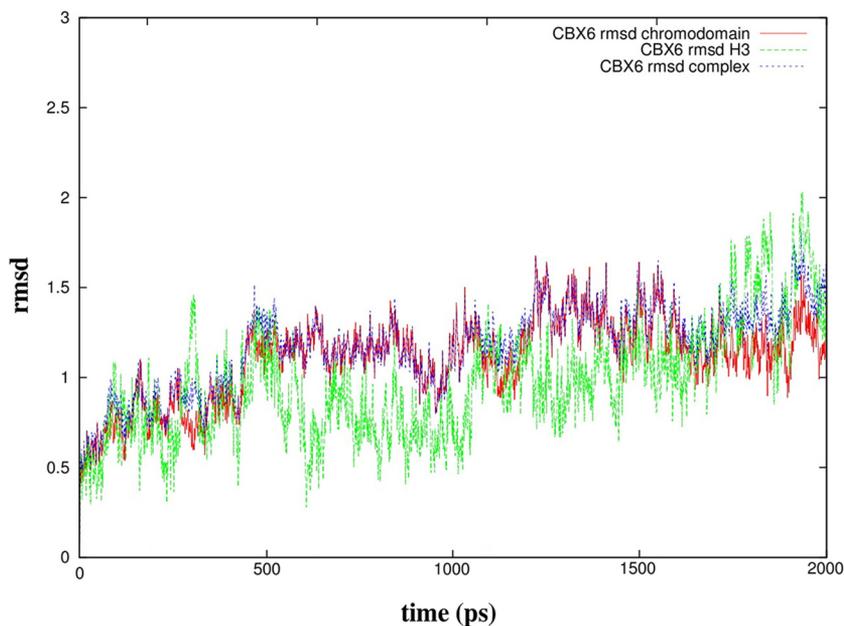
TABLE I  
Binding free energies of selected residues in the CBX6-H3K27me3 complex

Residue	Contribution (kcal/mol)	Residue	Contribution (kcal/mol)
Trp32	–4.77 ± 0.52	Glu46	–0.99 ± 0.28
Phe11	–3.78 ± 0.81	Arg9	–0.79 ± 0.94
Leu49	–3.62 ± 0.47	Glu58	–0.61 ± 0.04
Asn47	–3.40 ± 0.73	Val30	–0.51 ± 0.09
Asp50	–2.88 ± 0.52	Glu45	–0.45 ± 0.02
Val10	–2.77 ± 0.39	H3Q19	+0.08 ± 0.67
Arg52	–2.41 ± 1.14	H3L20	–0.72 ± 0.65
Leu53	–2.15 ± 0.30	H3A21	–0.45 ± 0.25
Thr41	–2.04 ± 0.32	H3T22	–1.53 ± 0.43
Ala13	–1.98 ± 0.43	H3K23	–6.20 ± 0.77
Trp35	–1.98 ± 0.60	H3A24	–4.64 ± 0.51
Glu43	–1.88 ± 0.74	H3A25	–6.05 ± 0.55
Ala12	–1.66 ± 0.24	H3R26	–7.22 ± 0.77
Glu14	–1.29 ± 0.37	H3K27me3	–11.79 ± 0.93
Ile48	–1.22 ± 0.22	H3S28	–1.73 ± 0.71
Tyr39	–1.04 ± 0.30	H3A29	–1.14 ± 0.53

through van der Waals attractions (Trp35, with a –1.98 kcal/mol contribution, is the third). Aside from H3K27me3 (–11.79 kcal/mol), the greatest contributions to the binding free energy from histone residues arise from H3R26 (–7.22 kcal/mol), H3K23 (–6.20 kcal/mol), H3A25 (–6.05 kcal/mol), and H3A24 (–4.64 kcal/mol). Trajectory visualization revealed that there are some important non-van der Waals interactions, such as the salt bridge between Glu46 and H3R26 and the hydrogen bond between Glu43 and H3S28. However, a survey of the residue contributions in Table I shows that van der Waals interactions are the main force in the binding between CBX6 and H3, accounting for 73% of the energy that is responsible for the complex binding. To further quantify the effects of the mutation of residues that are critical to the binding, we calculated the mutation matrix (Fig. 2, Supplementary Material) from *in silico* point mutation experiments (virtual mutagenesis) (73, 74) as a filter for possible CBX-binding peptides.

**Proteome-wide Identification of Methyllysine Peptides Binding to CBX6 Chromodomain**—CBX6 is a known reader of histone modifications through chromodomain-methyllysine recognition. We hypothesize that it also recognizes methyllysines in non-histone peptides sharing structural and energetic features with histone peptides. In order to identify these peptides in the human proteome, we developed a pipeline to integrate diverse data for a systematic search. We used the 11-residue-long H3K27me3 peptide as the template. Because the methyllysine is at the ninth position of the H3K27 sequence QLATKAARKSA, we required that the putative binding peptides of CBX6 contain a lysine at this position and have a binding energy with CBX6 comparable to that of H3K27me3. Because the number of candidate peptides was very large, we employed additional structural and functional criteria to reduce false positives (Fig. 3).

**FIG. 1. RMSD for the equilibration and production runs of CBX6 protein complexed with H3K27me3.** The red line is for the CBX6 protein, the green line for the H3 peptide, and the blue line for the entire complex.



**FIG. 2. Mutation matrix for the 10 residues in H3 apart from K27 that occur in the crystal structure used for the simulation.** Red cells indicate mutations relatively unfavorable for binding to CBX6 protein, and green cells show mutations that are favorable relative to wild-type H3.

We first extracted all 633,969 possible methyllysine-containing peptides (all 11-residue-long peptides containing a lysine at the ninth position) from the 20,280 human proteins documented in the Uniprot/Swissprot database (June 16, 2009) (92) (Fig. 3). If a peptide segment of a protein mediates the interaction with CBX6, the peptide and also the lysine residue to be methylated are under functional constraint and should be conserved during evolution. Therefore, we kept only the peptides that were conserved for the majority of the 11 positions (59,495 peptides passed the cutoff), and especially for the ninth lysine (39,631 peptides remained), across 42 vertebrate proteomes (supplemental Table S2).

Mass spectrometry data are considered as the gold standard for determining PTM. Because of the lack of effective enrichment methods for detecting methylated peptides, it is risky to rely only on the mass spectrometry data to define methyllysine, but a signal of methylation in a mass spectrometry study is still a good indicator that should be considered. Therefore, we searched for methylation signals of the remain-

ing peptides in 30 million mass spectra from human tissues. Given that these mass spectroscopy experiments were not designed to search for methyllysine, we included all 9878 peptides that showed signals of mono-, di-, or trimethylation on the lysine at the ninth position.

Next, we applied two filters that accounted for the structural features of the chromodomain binding peptides. Because in the template complex of CBX6-H3K27me3 the H3 peptide adopts an extended conformation, we removed any peptides that are observed to form a helix in crystal structure, if available, or are predicted to form a helix by secondary structure prediction methods (see “Materials and Methods”). About half of the peptides did not pass this filter, and 4153 remained. In addition, the putative binding peptides of the chromodomain should be exposed on the protein surface to form the complex. Therefore, if the structure of the protein containing the candidate peptide was available in the Protein Data Bank (60), we calculated the SASA for the peptide and kept only peptides with SASAs comparable to

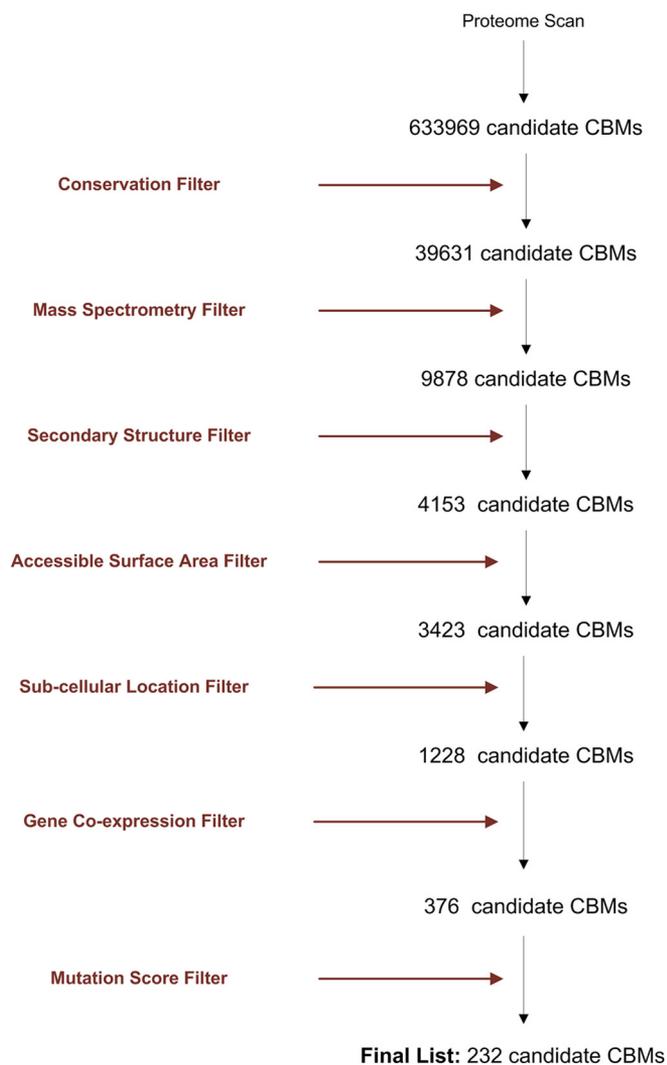


FIG. 3. Flowchart of the proteomic scan of methyllysine peptides binding to the CBX6 chromodomain.

that of the H3 peptide. As a result, 3423 peptides passed this filter.

To narrow down the candidate peptides for experimental investigation, we filtered the candidates using two functional criteria. Given that the CBX6 protein functions in the nucleus, it is likely that its interacting partners are also localized to the same cellular compartment. For the 1228 peptides that passed this localization filter, we calculated the correlation of gene expression profiles between CBX6 and the proteins containing the candidate peptides. Because the interacting partners are often co-expressed in various cell types or under multiple conditions, we kept only the proteins whose expression levels were strongly correlated with CBX6, which resulted in a list of 376 candidate peptides.

Lastly, the candidate peptides should bind to the CBX6 chromodomain with an affinity comparable to that of the H3 peptide. To prioritize peptides for experimental tests, we ranked the 376 peptides using their scores calculated from

the mutation matrix and removed 144 peptides with positive scores (weaker binders to the CBX6 chromodomain than the H3K27me3 peptide). The final 232 peptide sequences were printed on a peptide microarray for an *in vitro* binding assay (supplemental Table S3).

**Peptide Microarray Assay to Identify CBX6 Binders**—The chromodomain (residue positions 11–69) of human CBX6 protein was screened against a total of 248 nine-residue-long peptide sequences bearing one methylated lysine at the eighth position using a peptide microarray platform fabricated as described in Ref. 90. The 248 peptides consisted of the 232 selected non-histone methylated peptides, 6 methylated H3K9 and H3K27 peptides (mono-, di-, and trimethylated) as positive controls, and 10 unmodified SH3-domain-binding peptide sequences as negative controls (supplemental Table S3). All the peptides were immobilized onto the glass chip via an amine group on the C terminus followed by a hexa-carbon spacer to provide more separation and flexibility of the peptide strand relative to the surface. The CBX6 chromodomain was incubated with the printed peptides on the chip under constant agitation to allow sufficient contact. Binding between the protein and any peptide was detected by visualizing the immobilized protein on the chip with a fluorescent-labeled antibody.

We validated the proper functioning of the current peptide microarray platform and the analysis criteria for CBX6 chromodomain binding using the positive and negative controls. Although no negative control peptides showed any binding activity to CBX6 chromodomain on the microarray, the di- and trimethylated H3K9 and H3K27 peptides were identified as strong binders, with the trimethylated H3K27me3 showing the highest signal intensity. Monomethylated H3K9me and H3K27me both showed fluorescence intensity around the binder/non-binder boundary (with only one out of the three signals above the cutoff value), and thus these signals were considered below the confidence threshold. Given that previous solution binding studies (29) reported a weak binding affinity of these peptides toward CBX6 chromodomain ( $K_D > 500 \mu\text{M}$  for both trimethylated H3K9me3 and H3K27me3) despite the crystal structures for both complexes achieved, our observations suggested a superior detection sensitivity exhibited by this peptide microarray platform relative to conventional solution methods. To confirm the legitimacy of these weak binding interactions detected on the microarray, additional pull-down experiments were conducted for CBX6 chromodomain with the methylated H3K9 and H3K27 peptides. All di- and trimethylated peptide coated resins successfully immobilized the CBX6 chromodomain in the pull-down experiments, with the trimethylated H3K27me3 showing the highest binding affinity (supplemental Fig. S1), consistent with the microarray results.

A gradient of fluorescent intensities from the protein-incubated peptides was observed throughout the microarray, and a statistical cutoff was selected to quantitatively distinguish

## Identification of Methyllysine Peptides Binding to CBX6 Chromodomain in the Human Proteome

the binders from the non-binders. For each peptide, the fluorescent intensity of the printed triplet was individually measured and fit to a mixed Gaussian distribution with two components (non-binding background *versus* binder) (Fig. 4). A total of 50 non-histone peptides (Table II) were identified as putative binders for the CBX6 chromodomain with significant

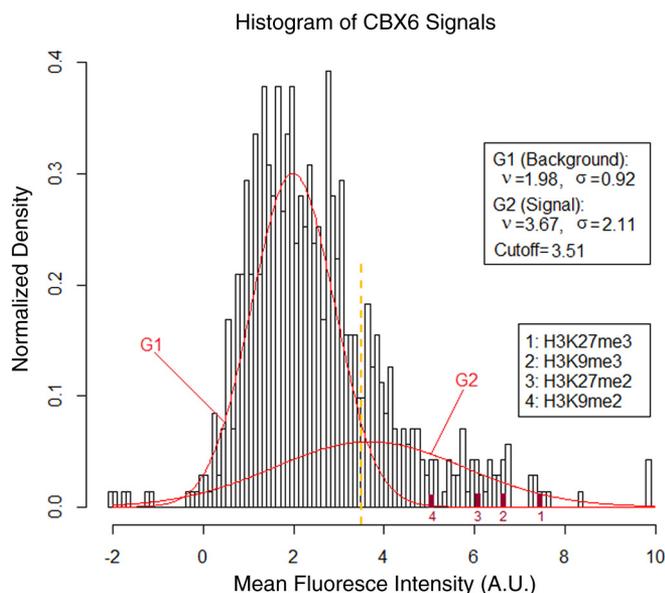


FIG. 4. Gaussian mixture fitting of signal intensities from peptide microarray for the CBX6-GST fusion protein.

fluorescence intensities above a statistical cutoff of  $p < 0.05$  (from the background Gaussian distribution). Given that most of the peptides passing the computational filters might be possible binders to the CBX6 chromodomain, the cutoff selected based on the distribution of these peptides is likely a stringent one that missed some true positives.

**Functional Analysis of the Putative Binders**—As the interacting partners of polycomb family proteins are still poorly defined, the newly identified putative binding peptides of CBX6 would reveal unknown functions of the chromatin remodeling complexes and their interplay with other important cellular regulators. We input the proteins that contained the binding peptides from the microarray experiments into DAVID (93), a tool for function annotation clustering analysis using Gene Ontology terms. We found that these proteins were functionally clustered into six groups (cluster cutoff: enrichment score  $> 1.3$ ; in each cluster, only Gene Ontology terms with  $p$  values  $< 0.05$  are shown in supplemental Table S4).

CBX6 is able to recognize histone repression marks H3K9me3 and H3K27me3 (27–30). CBX6 is a component of PRC1, which can alter chromatin structure and repress gene transcription. It is not surprising that a significant portion of array-identified CBX6-binding proteins play roles in chromatin modification. For example, NSD1 and ASH1L are two histone methyltransferases with different substrate specificities. NSD1 methylates H3K36 and H4K20, and ASH1L specifically methylates H3K36. These marks are usually found in the gene

TABLE II  
 CBX6 binders identified by microarray

Uniprot ID	Sequence <sup>a</sup>	Uniprot ID	Sequence
ZBT24_HUMAN	ECKDCHR[K_Me3]F	ISCU_HUMAN	ELCLPPV[K_Me3]L
TLE3_HUMAN	DISHDGT[K_Me3]L	MYST3_HUMAN	SSCRDQG[K_Me3]N
NSD1_HUMAN	IHTCFVC[K_Me3]Q	TLX2_HUMAN	YQNRTPP[K_Me3]R
CHD8_HUMAN	EEDDEGS[K_Me3]F	MYST4_HUMAN	CSFCLGT[K_Me3]E
ROA3_HUMAN	PGAHLTV[K_Me3]K	ADNP_HUMAN	EQAVYYC[K_Me3]K
ZF64B_HUMAN	HSDERPY[K_Me3]C	KDM2B_HUMAN	PNCWECP[K_Me3]C
INT6_HUMAN	CEPMVID[K_Me3]L	ASH1L_HUMAN	GLMIQCD[K_Me3]C
ZN652_HUMAN	SCNKSFK[K_Me3]L	TAF5L_HUMAN	LAFPSNG[K_Me3]Y
CTCF_HUMAN	HTDERPH[K_Me3]C	ETV3_HUMAN	ILHKTGK[K_Me3]R
TRI33_HUMAN	PVFCPVH[K_Me3]Q	PLAL2_HUMAN	HQCMYCD[K_Me3]M
ZBT47_HUMAN	IVHGYAE[K_Me3]K	SMCE1_HUMAN	YDDGFSM[K_Me3]H
RELB_HUMAN	ELYLLCD[K_Me3]V	ASXL2_HUMAN	DDCIGPS[K_Me3]L
TET1_HUMAN	SHGCPIA[K_Me3]W	SIRT1_HUMAN	TASCLIC[K_Me3]Y
FGF14_HUMAN	PAAHFLP[K_Me3]P	ZN668_HUMAN	YSCERC[G_K_Me3]A
ZN627_HUMAN	YKCKVCG[K_Me3]A	CDYL2_HUMAN	KRKNKKG[K_Me3]W
BUD31_HUMAN	ETEPHEG[K_Me3]R	ZN768_HUMAN	YKCPHCG[K_Me3]A
AGO2_HUMAN	QGQPCFC[K_Me3]Y	CTCFL_HUMAN	LTGKPF[K_Me3]C
ZBT41_HUMAN	HQCDVCK[K_Me3]I	KKCC1_HUMAN	HGDDGHV[K_Me3]I
RORB_HUMAN	NTVLFEG[K_Me3]Y	ZN772_HUMAN	YKCSECG[K_Me3]A
RBBP5_HUMAN	VHPLLGV[K_Me3]G	ZNF35_HUMAN	FACNDCG[K_Me3]A
ZN345_HUMAN	YVCKECG[K_Me3]A	BRD1_HUMAN	GACIQCH[K_Me3]A
RBM39_HUMAN	AQGNVYV[K_Me3]C	ZRAB2_HUMAN	EDDADLS[K_Me3]Y
PSA7_HUMAN	SGTYHAW[K_Me3]A	MED1_HUMAN	GKSPSRN[K_Me3]K
PRP4_HUMAN	CSYDRTF[K_Me3]L	ZN184_HUMAN	YKNECG[K_Me3]A
SIN3A_HUMAN	MSDENC[F_K_Me3]L	SFRS1_HUMAN	VRKLDNT[K_Me3]F

<sup>a</sup> [K\_Me3] stands for trimethylated-lysine residue.

body of transcribed genes. We also identified histone acetyltransferases MYST3 and MYST4 that may interact with CBX6. MYST3 and MYST4 are components of the MOZ/MORF complex, which preferentially acetylates H3. In addition, histone demethylase KDM2B, which removes methylation of H3K4 and H3K36, and histone deacetylase SIRT1, which represses gene expression by deacetylating H3K9ac, also potentially interact with CBX6. Although the question of whether these interactions occur *in vivo* has yet to be addressed, it is very reasonable to speculate that CBX6 mediates cross-talk between different histone-modifying enzymes.

We also found proteins in the putative binder list that play critical roles in chromatin remodeling and transcriptional regulation. SMCE1 is a component of the BAF complex and regulates gene expression via chromatin remodeling. CDYL2 and CHD8 both contain chromodomain and repress the expression of their target genes. CTCF is well known as an insulator of the genome and blocks interaction between enhancers and promoters, preventing the spread of heterochromatin. Moreover, several proteins involved in various gene regulation and development processes besides chromatin remodeling were also pinpointed. AGO2 is required for RNA-mediated gene silencing by the RNA-induced silencing complex, MED1 is a component of Mediator complex co-activators of nearly all Pol II-dependent genes, TET1 is a dioxygenase that plays the important role of hydroxylating methylated cytosine and is involved in embryonic stem cell maintenance, RELB is a component of NF- $\kappa$ B, SIN3A represses Myc-related genes, and TLE3 inhibits the Wnt signaling pathway.

In addition, the proteins KKCC1 and SFRS1 are in our array-positive list and have also been identified by Levy *et al.* (31) as substrates of the lysine methyltransferase SETD6. KKCC1 is a calcium/calmodulin-dependent kinase involved in the regulation of cell apoptosis, and SFRS1 is related to exon skipping, a mechanism that ensures correct splicing. Taken as a whole, our study not only identifies putative methyllysine in the human proteome, but also suggests the functional consequences of this PTM in mediating protein–protein interaction and signal transduction.

#### CONCLUSION AND DISCUSSION

We present here a systematic approach that integrates structural modeling, bioinformatics analysis, and peptide microarray binding assay to identify methyllysine-containing peptides in the human proteome that may bind to the CBX6 chromodomain. The success of this approach with CBX6 suggests its potential broad application to other chromodomains or domains recognizing methylated lysine or arginine. Considering the difficulty of defining PTMs of methylation in proteomic studies, this proof-of-concept study opens a new avenue for defining methylation events.

The newly identified interacting partners of CBX6 reveal many possible cross-talks between different regulatory mech-

anisms, including histone modification, chromatin remodeling, DNA methylation, RNA splicing, and microRNA processing. For example, previous epigenomic studies have identified bivalent promoters (94) and enhancers (95, 96)—that is, the promoters/enhancers are “poised” and marked by both active (e.g. H3K4me3 for promoter and H3K4me1/H3K27ac for enhancer) and repressive (H3K27me3 for both promoter and enhancer) histone modifications. The promoters and enhancers can become active or remain silenced, depending upon cellular conditions. CBX6 is a component of the PRC1 complex, which collaborates with PRC2 to silence genes through H3K27me3 (97). Therefore, we speculate that the putative interactions identified by our study might imply cross-talk between multiple chromatin remodeling complexes and that CBX6’s interactions with both active (e.g. NSD1, ASH1L, MYST3/4) and repressive (e.g. KDM2B, SIRT1) factors might make it possible for the mechanisms underlying regulation of the poised promoters/enhancers to be deciphered. Although these findings are still waiting for *in vivo* confirmation, the specific predictions of the methylation site and interaction between CBX6 and other proteins would greatly facilitate further investigation of the biological significance.

Although peptide microarray technology has been widely exploited in the study of protein–peptide interactions, its combination with computational screening has proven to be important in proteomic studies, because the number of candidate peptides is normally huge and it is financially prohibitive to synthesize all of them (90). In this and our previous studies (90), bioinformatics analyses that integrate diverse and complementary information such as binding affinity estimated by mutation matrix and conservation across species have greatly boosted the hit rate of peptide microarrays, allowing us to find binding peptides from the human proteome. We thus believe such an integrated approach is particularly useful in identifying PTMs that are difficult to define otherwise and revealing how these PTMs regulate protein functions.

*Acknowledgments*—Some of the molecular dynamics simulations were conducted in the Center for Theoretical Biological Physics (CTBP) computing cluster at the University of California, San Diego.

\* This work was partially supported by the National Institutes of Health (GM085188 to W.W.).

§ This article contains [supplemental material](#).

|| To whom correspondence should be addressed: E-mail: weiwang@ucsd.edu.

§ These authors contributed to this work equally.

#### REFERENCES

- Walsh, C. T., Garneau-Tsodikova, S., and Gatto, G. J., Jr. (2005) Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem. Int. Ed. Engl.* **44**, 7342–7372
- Tan, C. S. (2011) Sequence, structure, and network evolution of protein phosphorylation. *Sci. Signal.* **4**, mr6
- Boscher, C., Dennis, J. W., and Nabi, I. R. (2011) Glycosylation, galectins and cellular signaling. *Curr. Opin. Cell Biol.* **23**, 383–392
- Soppa, J. (2010) Protein acetylation in archaea, bacteria, and eukaryotes. *Archaea* **2010**

5. Jones, J. D., and O'Connor, C. D. (2011) Protein acetylation in prokaryotes. *Proteomics* **11**, 3012–3022
6. Shi, Y., Xu, P., and Qin, J. (2011) Ubiquitinated proteome: ready for global? *Mol. Cell. Proteomics* **10**, R110.006882
7. Hannoun, Z., Greenhough, S., Jaffray, E., Hay, R. T., and Hay, D. C. (2010) Post-translational modification by SUMO. *Toxicology* **278**, 288–293
8. Wilkinson, K. A., and Henley, J. M. (2010) Mechanisms, regulation and consequences of protein SUMOylation. *Biochem. J.* **428**, 133–145
9. Suganuma, T., and Workman, J. L. (2011) Signals and combinatorial functions of histone modifications. *Annu. Rev. Biochem.* **80**, 473–499
10. Bedford, M. T., and Clarke, S. G. (2009) Protein arginine methylation in mammals: who, what, and why. *Mol. Cell* **33**, 1–13
11. Pradhan, S., Chin, H. G., Esteve, P. O., and Jacobsen, S. E. (2009) SET7/9 mediated methylation of non-histone proteins in mammalian cells. *Epigenetics* **4**, 383–387
12. Spange, S., Wagner, T., Heinzel, T., and Kramer, O. H. (2009) Acetylation of non-histone proteins modulates cellular signalling at multiple levels. *Int. J. Biochem. Cell Biol.* **41**, 185–198
13. Ruthenburg, A. J., Li, H., Patel, D. J., and Allis, C. D. (2007) Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* **8**, 983–994
14. Kouzarides, T. (2007) Chromatin modifications and their function. *Cell* **128**, 693–705
15. Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., Lu, Z., Ye, Z., Zhu, Q., Wysocka, J., Ye, Y., Khochbin, S., Ren, B., and Zhao, Y. (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146**, 1016–1028
16. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., and Patel, D. J. (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **14**, 1025–1040
17. Yun, M., Wu, J., Workman, J. L., and Li, B. (2011) Readers of histone modifications. *Cell Res.* **21**, 564–578
18. Kerppola, T. K. (2009) Polycomb group complexes—many combinations, many functions. *Trends Cell Biol.* **19**, 692–704
19. Morey, L., and Helin, K. (2010) Polycomb group protein-mediated repression of transcription. *Trends Biochem. Sci.* **35**, 323–332
20. 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. U.S.A.* **88**, 263–267
21. Martinez, A. M., and Cavalli, G. (2006) The role of polycomb group proteins in cell cycle regulation during development. *Cell Cycle* **5**, 1189–1197
22. Sparmann, A., and van Lohuizen, M. (2006) Polycomb silencers control cell fate, development and cancer. *Nat. Rev. Cancer* **6**, 846–856
23. Rajasekhar, V. K., and Begemann, M. (2007) Concise review: roles of polycomb group proteins in development and disease: a stem cell perspective. *Stem Cells* **25**, 2498–2510
24. Wang, J., Mager, J., Chen, Y., Schneider, E., Cross, J. C., Nagy, A., and Magnuson, T. (2001) Imprinted X inactivation maintained by a mouse Polycomb group gene. *Nat. Genet.* **28**, 371–375
25. de Napoles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., Nesterova, T. B., Silva, J., Otte, A. P., Vidal, M., Koseki, H., and Brockdorff, N. (2004) Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev. Cell* **7**, 663–676
26. Vandamme, J., Volkel, P., Rosnoblet, C., Le Faou, P., and Angrand, P. O. (2011) Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells. *Mol. Cell. Proteomics* **10**, M110.002642
27. Min, J., Zhang, Y., and Xu, R. M. (2003) Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* **17**, 1823–1828
28. Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D., and Khorasanizadeh, S. (2003) Molecular basis for the discrimination of repressive methyllysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **17**, 1870–1881
29. Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., Wasney, G. A., Li, Z., Vedadi, M., Schapira, M., Min, J., and Arrowsmith, C. H. (2011) Recognition and specificity determinants of the human CBX chromodomains. *J. Biol. Chem.* **286**, 521–529
30. Yap, K. L., Li, S., Munoz-Cabello, A. M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M. J., and Zhou, M. M. (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol. Cell* **38**, 662–674
31. Levy, D., Liu, C. L., Yang, Z., Newman, A. M., Alizadeh, A. A., Utz, P. J., and Gozani, O. (2011) A proteomic approach for the identification of novel lysine methyltransferase substrates. *Epigenetics Chromatin* **4**, 19
32. Dhayalan, A., Kudithipudi, S., Rathert, P., and Jeltsch, A. (2011) Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase. *Chem. Biol.* **18**, 111–120
33. Balasubramanian, N., Ananthanarayanan, M., and Suchy, F. J. (2012) Direct methylation of FXR by Set7/9, a lysine methyltransferase, regulates the expression of FXR target genes. *Am. J. Physiol. Gastrointest. Liver Physiol.* **302**, G937–G947
34. Huang, J., Perez-Burgos, L., Placek, B. J., Sengupta, R., Richter, M., Dorsey, J. A., Kubicek, S., Opravil, S., Jenuwein, T., and Berger, S. L. (2006) Repression of p53 activity by Smyd2-mediated methylation. *Nature* **444**, 629–632
35. Chuikov, S., Kurash, J. K., Wilson, J. R., Xiao, B., Justin, N., Ivanov, G. S., McKinney, K., Tempst, P., Prives, C., Gambelin, S. J., Barlev, N. A., and Reinberg, D. (2004) Regulation of p53 activity through lysine methylation. *Nature* **432**, 353–360
36. Couture, J. F., Collazo, E., Hauk, G., and Trievel, R. C. (2006) Structural basis for the methylation site specificity of SET7/9. *Nat. Struct. Mol. Biol.* **13**, 140–146
37. Huang, J., Dorsey, J., Chuikov, S., Perez-Burgos, L., Zhang, X., Jenuwein, T., Reinberg, D., and Berger, S. L. (2010) G9a and Glp methylate lysine 373 in the tumor suppressor p53. *J. Biol. Chem.* **285**, 9636–9641
38. Shi, X., Kachirskaja, I., Yamaguchi, H., West, L. E., Wen, H., Wang, E. W., Dutta, S., Appella, E., and Gozani, O. (2007) Modulation of p53 function by SET8-mediated methylation at lysine 382. *Mol. Cell* **27**, 636–646
39. West, L. E., Roy, S., Lachmi-Weiner, K., Hayashi, R., Shi, X., Appella, E., Kutateladze, T. G., and Gozani, O. (2010) The MBT repeats of L3MBTL1 link SET8-mediated p53 methylation at lysine 382 to target gene repression. *J. Biol. Chem.* **285**, 37725–37732
40. Wang, J., Hevi, S., Kurash, J. K., Lei, H., Gay, F., Bajko, J., Su, H., Sun, W., Chang, H., Xu, G., Gaudet, F., Li, E., and Chen, T. (2009) The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat. Genet.* **41**, 125–129
41. Esteve, P. O., Chin, H. G., Benner, J., Feehery, G. R., Samaranyake, M., Horwitz, G. A., Jacobsen, S. E., and Pradhan, S. (2009) Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5076–5081
42. Ghoshal, K., Datta, J., Majumder, S., Bai, S., Kutay, H., Motiwala, T., and Jacob, S. T. (2005) 5-Aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN box, bromo-adjacent homology domain, and nuclear localization signal. *Mol. Cell. Biol.* **25**, 4727–4741
43. Levy, D., Kuo, A. J., Chang, Y., Schaefer, U., Kitson, C., Cheung, P., Espejo, A., Zee, B. M., Liu, C. L., Tangsombatvisit, S., Tennen, R. I., Kuo, A. Y., Tanjing, S., Cheung, R., Chua, K. F., Utz, P. J., Shi, X., Prinjha, R. K., Lee, K., Garcia, B. A., Bedford, M. T., Tarakhovskiy, A., Cheng, X., and Gozani, O. (2011) Lysine methylation of the NF-kappaB subunit RelA by SETD6 couples activity of the histone methyltransferase GLP at chromatin to tonic repression of NF-kappaB signaling. *Nat. Immunol.* **12**, 29–36
44. Wunderlich, Z., and Mirny, L. A. (2009) Using genome-wide measurements for computational prediction of SH2-peptide interactions. *Nucleic Acids Res.* **37**, 4629–4641
45. Hou, T., Zhang, W., Case, D. A., and Wang, W. (2008) Characterization of domain-peptide interaction interface: a case study on the amphiphysin-1 SH3 domain. *J. Mol. Biol.* **376**, 1201–1214
46. Hou, T., Xu, Z., Zhang, W., McLaughlin, W. A., Case, D. A., Xu, Y., and Wang, W. (2009) Characterization of domain-peptide interaction interface: a generic structure-based model to decipher the binding specificity of SH3 domains. *Mol. Cell. Proteomics* **8**, 639–649
47. Zhang, L., Shao, C., Zheng, D., and Gao, Y. (2006) An integrated machine learning system to computationally screen protein databases for protein binding peptide ligands. *Mol. Cell. Proteomics* **5**, 1224–1232
48. Ferraro, E., Via, A., Ausiello, G., and Helmer-Citterich, M. (2006) A novel structure-based encoding for machine-learning applied to the inference of SH3 domain specificity. *Bioinformatics* **22**, 2333–2339
49. Brannetti, B., Via, A., Cestra, G., Cesareni, G., and Helmer-Citterich, M.

- (2000) SH3-SPOT: an algorithm to predict preferred ligands to different members of the SH3 gene family. *J. Mol. Biol.* **298**, 313–328
50. Stiffler, M. A., Chen, J. R., Grantcharova, V. P., Lei, Y., Fuchs, D., Allen, J. E., Zaslavskaya, L. A., and MacBeath, G. (2007) PDZ domain binding selectivity is optimized across the mouse proteome. *Science* **317**, 364–369
  51. Chen, J. R., Chang, B. H., Allen, J. E., Stiffler, M. A., and MacBeath, G. (2008) Predicting PDZ domain-peptide interactions from primary sequences. *Nat. Biotechnol.* **26**, 1041–1045
  52. Smith, C. A., and Kortemme, T. (2010) Structure-based prediction of the peptide sequence space recognized by natural and synthetic PDZ domains. *J. Mol. Biol.* **402**, 460–474
  53. Kalyoncu, S., Keskin, O., and Gursoy, A. (2010) Interaction prediction and classification of PDZ domains. *BMC Bioinformatics* **11**, 357
  54. Hui, S., and Bader, G. D. (2010) Proteome scanning to predict PDZ domain interactions using support vector machines. *BMC Bioinformatics* **11**, 507
  55. Li, N., Hou, T., Ding, B., and Wang, W. (2011) Characterization of PDZ domain-peptide interaction interface based on energetic patterns. *Proteins* **79**, 3208–3220
  56. Chen, H., Xue, Y., Huang, N., Yao, X., and Sun, Z. (2006) MeMo: a web tool for prediction of protein methylation modifications. *Nucleic Acids Res.* **34**, W249–W253
  57. Shien, D. M., Lee, T. Y., Chang, W. C., Hsu, J. B., Horng, J. T., Hsu, P. C., Wang, T. Y., and Huang, H. D. (2009) Incorporating structural characteristics for identification of protein methylation sites. *J. Comput. Chem.* **30**, 1532–1543
  58. Shao, J., Xu, D., Tsai, S. N., Wang, Y., and Ngai, S. M. (2009) Computational identification of protein methylation sites through bi-profile Bayes feature extraction. *PLoS One* **4**, e4920
  59. Hu, L. L., Li, Z., Wang, K., Niu, S., Shi, X. H., Cai, Y. D., and Li, H. P. (2011) Prediction and analysis of protein methylarginine and methyllysine based on Multisequence features. *Biopolymers* **95**, 763–771
  60. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242
  61. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688
  62. Frisch, M. J., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J. A., Jr., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., and Pople, J. A. (2004) *Gaussian 03, Revision C*, 2nd ed., Gaussian Inc., Wallingford, CT
  63. Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges—the RESP model. *J. Phys. Chem.* **97**, 10269–10280
  64. Cieplak, P., Cornell, W. D., Bayly, C., and Kollman, P. A. (1995) Application of the multimolecule and multiconformational RESP methodology to biopolymers—charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **16**, 1357–1377
  65. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G. M., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J. M., and Kollman, P. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012
  66. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174
  67. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935
  68. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690
  69. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-integration of Cartesian equations of motion of a system with constraints—molecular-dynamics of N-alkanes. *J. Comput. Phys.* **23**, 327–341
  70. Xiang, Z., and Honig, B. (2001) Extending the accuracy limits of prediction for side-chain conformations. *J. Mol. Biol.* **311**, 421–430
  71. Wang, W., Donini, O., Reyes, C. M., and Kollman, P. A. (2001) Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid non-covalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 211–243
  72. Hou, T., McLaughlin, W. A., and Wang, W. (2008) Evaluating the potency of HIV-1 protease drugs to combat resistance. *Proteins* **71**, 1163–1174
  73. Hou, T., Chen, K., McLaughlin, W. A., Lu, B., and Wang, W. (2006) Computational analysis and prediction of the binding motif and protein interacting partners of the Abl SH3 domain. *PLoS Comput. Biol.* **2**, e1
  74. Wang, W., and Kollman, P. A. (2000) Free energy calculations on dimer stability of the HIV protease using molecular dynamics and continuum solvent model. *J. Mol. Biol.* **303**, 567–582
  75. Onufriev, A., Bashford, D., and Case, D. A. (2000) Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **104**, 3712–3720
  76. Weiser, J., Shenkin, P. S., and Still, W. C. (1999) Approximate solvent-accessible surface areas from tetrahedrally directed neighbor densities. *Biopolymers* **50**, 373–380
  77. Pellegrini, E., and Field, M. J. (2002) A generalized-Born solvation model for macromolecular hybrid-potential calculations. *J. Phys. Chem. A* **106**, 1316–1326
  78. Hubbard, T. J., Aken, B. L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Clarke, L., Coates, G., Fairley, S., Fitzgerald, S., Fernandez-Banet, J., Gordon, L., Graf, S., Haider, S., Hammond, M., Holland, R., Howe, K., Jenkinson, A., Johnson, N., Kahari, A., Keefe, D., Keenan, S., Kinsella, R., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Megy, K., Meidl, P., Overduin, B., Parker, A., Pritchard, B., Rios, D., Schuster, M., Slater, G., Smedley, D., Spooner, W., Spudich, G., Trevanion, S., Vilella, A., Vogel, J., White, S., Wilder, S., Zadissa, A., Birney, E., Cunningham, F., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Herrero, J., Kasprzyk, A., Proctor, G., Smith, J., Searle, S., and Flicek, P. (2009) Ensembl 2009. *Nucleic Acids Res.* **37**, D690–D697
  79. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410
  80. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948
  81. Henikoff, S., and Henikoff, J. G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10915–10919
  82. Vizcaino, J. A., Cote, R., Reisinger, F., Foster, J. M., Mueller, M., Rameseder, J., Hermjakob, H., and Martens, L. (2009) A guide to the Proteomics Identifications Database proteomics data repository. *Proteomics* **9**, 4276–4283
  83. Deutsch, E. W., Lam, H., and Aebersold, R. (2008) PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep.* **9**, 429–434
  84. Craig, R., and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467
  85. Tanner, S., Shu, H., Frank, A., Wang, L. C., Zandi, E., Mumby, M., Pevzner, P. A., and Bafna, V. (2005) InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal. Chem.* **77**, 4626–4639
  86. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637
  87. Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202

88. Cheng, J., Randall, A. Z., Sweredoski, M. J., and Baldi, P. (2005) SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res.* **33**, W72–W76
89. Hubble, J., Demeter, J., Jin, H., Mao, M., Nitzberg, M., Reddy, T. B., Wymore, F., Zachariah, Z. K., Sherlock, G., and Ball, C. A. (2009) Implementation of GenePattern within the Stanford Microarray Database. *Nucleic Acids Res.* **37**, D898–D901
90. Xu, Z., Hou, T., Li, N., Xu, Y., and Wang, W. (2012) Proteome-wide detection of Abl1 SH3-binding peptides by integrating computational prediction and peptide microarray. *Mol. Cell. Proteomics* **11**, O111.010389
91. Rasband, W. S. (1997–2008) *ImageJ*, National Institutes of Health, Bethesda, MD
92. UniProt Consortium (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.* **38**, D142–D148
93. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57
94. Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., and Lander, E. S. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326
95. Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., and Wysocka, J. (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283
96. Creighton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., and Jaenisch, R. (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21931–21936
97. Simon, J. A., and Kingston, R. E. (2009) Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* **10**, 697–708