Characterization of structural variability sheds light on the specificity determinants of the interaction between effector domains and histone tails

Sergi Lois, Naiara Akizu, Gemma Mas de Xaxars, Iago Vázquez, Marian Martínez-Balbás & Xavier de la Cruz

To cite this article: Sergi Lois, Naiara Akizu, Gemma Mas de Xaxars, Iago Vázquez, Marian Martínez-Balbás & Xavier de la Cruz (2010) Characterization of structural variability sheds light on the specificity determinants of the interaction between effector domains and histone tails, Epigenetics, 5:2, 137-148, DOI: 10.4161/epi.5.2.11079

To link to this article: http://dx.doi.org/10.4161/epi.5.2.11079

Published online: 16 Feb 2010.

Submit your article to this journal

Article views: 33

View related articles

Citing articles: 5 View citing articles
**Characterization of structural variability sheds light on the specificity determinants of the interaction between effector domains and histone tails**

Sergi Lois,1,2 Naiara Akizu,3 Gemma Mas de Xaxars,2 Iago Vázquez,2 Marian Martínez-Balbás3 and Xavier de la Cruz2,4,*

1Institut de Medicina Preventiva i Personalitzada del Cancer; Badalona, Spain; 2Computational Biology and Bioinformatics laboratory; and 3Molecular Signaling and Chromatin Function laboratory; Instituto de Biología Molecular de Barcelona; Consejo Superior de Investigaciones Científicas; Parc Científic de Barcelona; Barcelona, Spain; 4Institució Catalana de Recerca i Estudis Avançats; Barcelona, Spain

**Key words:** histone-effector interaction, modification of histone tails, histone binding mode, epigenetic regulator specificity, histone disorder, lysine post-translational modifications

**Abbreviations:** BS, whole binding site; MBS, binding site of the modified residue; AUTODOM, comparisons between BS from different versions of the same structure (experimental replicas, or different binding states, etc.); INTRADOM, INTERDOM, comparisons between BS from effectors with the same fold or different folds, respectively; ASA, accessible surface area

Introduction

Post-translational modifications—acetylation, methylation, ubiquitination, etc.—of the histone tails play a fundamental role in the regulation of gene expression by altering the chromatin state to either active or repressed.1,2 These marks are recognized by a set of effector modules able to bind the modified amino acids and part of their neighboring residues.1,2 Consequently, the structural characterization of the histone-effector interaction constitutes an important step towards understanding the molecular mechanisms of epigenetic regulation of gene expression, and how their failure leads to disease.

In recent years many studies have addressed this goal2 and there is now available a significant number of effector structures in either their apo- and/or holo-forms (where the bound compounds are usually histone peptides carrying different marks). The results of these studies have shed light on the structural transition experienced by histone tails upon effector binding, showing that it may vary depending on the local properties of the sequence stretch considered, thus allowing us to identify an additional specificity determinant for this interaction.

Overall, the results of our analysis contribute to clarify the origins of specificity: different regions of the binding site and, in particular, differences in the disorder-order transitions experienced by different histone sequence stretches upon binding.

In recent years many studies have addressed this goal2 and there is now available a significant number of effector structures in either their apo- and/or holo-forms (where the bound compounds are usually histone peptides carrying different marks). The results of these studies have shed light on the structural transition experienced by histone tails upon effector binding, showing that it may vary depending on the local properties of the sequence stretch considered, thus allowing us to identify an additional specificity determinant for this interaction.

Overall, the results of our analysis contribute to clarify the origins of specificity: different regions of the binding site and, in particular, differences in the disorder-order transitions experienced by different histone sequence stretches upon binding.
interactions and steric restraints playing a fundamental role.\textsuperscript{2} Considering together the known structures of histone-effector complexes, Taverna and colleagues\textsuperscript{2} have identified two main binding modes for marked lysines related to binding specificity: cavity-insertion and surface-groove. In the cavity-insertion mode, the modified lysine side chain is buried in a deep, narrow pocket with the ability to filter out ligands by size; on the contrary, in the surface-groove mode the lysine side chain is not subjected to such strict restraints. Histone peptides are intrinsically disordered;\textsuperscript{4} however, when bound to effectors they adopt an extended conformation.\textsuperscript{2} On this basis it has been proposed that the conformational transition experienced by histone tails between unbound and bound states corresponds to a disorder-order transition,\textsuperscript{4} rather than to an induced-fit mechanism.\textsuperscript{2} This is important because binding energetics depends on whether structural transitions are involved, and their kind.

The above findings clarify our view of the histone-effector interaction and point to some important open issues about the structural properties of the histone-effector interaction, and the corresponding biophysical/biochemical consequences. For example, our knowledge of specificity determinants is still incomplete. Data from different authors\textsuperscript{2–5} suggest that apart from the binding site of the modified residue (to which we will refer as MBS from now on) the remainder of the binding site may also play a role. Therefore, classifying whole binding sites (to which we will refer as BS from now on), and relating the resulting classes with the known MBS classes would allow a better understanding of how binding specificity is distributed over the different BS components. In addition, although it will not be considered here, good classifications of BS (1) can be used to improve comparative modeling and docking studies of histone-effector complexes, facilitating the identification of BS and binding modes in domains for which no structural information is available or is restricted to their apo forms, (2) allow a better understanding of the evolution of function mechanisms within protein families,\textsuperscript{5–7} and (3) allow the identification of possible sources of cross-reactions in designed ligands.\textsuperscript{8} In addition, to improve our knowledge of specificity determinants it is also important to understand how binding affects the histone peptide structure. For example, the degree to which the extended conformations in the different complexes are similar (it is known that extended conformations are structurally heterogeneous\textsuperscript{4}), if there are conserved structural motifs indicating the existence of structural propensities that could favor/disfavor the unbound-bound transition, whether this transition is the same for all modified lysines (e.g., H3K4, H3K9, H3K27, etc.). Identification of any structural trend in histone peptides would help to improve our quantitative understanding of the disorder-order transition associated to histone binding and to see if histone modifications owe part of their functional effect to shifts in this transition.

In this article we present the results of our work on two of these issues: the structural characterization of the BS of effectors, and that of the bound histone peptides. The BS of a series of effectors were compared using both visual and automatic structure comparison methods; peptide structures were also compared following a similar approach. Our results showed that BS could be classified in three main structural classes: cavity-insertion, flat-groove and narrow-groove. These classes did not completely coincide with those identified for the MBS\textsuperscript{2} pointing to a partition of the BS in specificity determinants. Regarding histone peptides, our results indicated that apart from a certain amount of structural heterogeneity, almost all of them shared a small, hook-like motif involving the side chain atoms of the modified lysine and some of the nearby main chain atoms. The different sequence propensities for the ARK and RTK sequences corresponding to this motif show that these tripeptides undergo a different disorder-order transition upon binding that constitutes a specificity determinant for the histone-effector interaction.

**Results**

**Visual analysis of the BS.** Visual analysis was done using a molecular surface\textsuperscript{10} representation of the BS, as it averages out atomic detail thus providing a global-shape view which allowed the classification of BS. Inspection of the available complexes (Table 1) showed that BS could be classified in three classes: flat-groove, narrow-groove and cavity-insertion. We found that there was not an exact mapping between these classes and Taverna and colleagues\textsuperscript{2} classes. Our cavity-insertion class was essentially the same as theirs. However, their MBS surface-groove class mapped to our two remaining BS classes: the flat-groove (Fig. 1) and narrow-groove (Fig. 2). The former included BS from the PHD, the Tudor and the double chromodomain effectors. The narrow-groove class was constituted only by BS from single chromodomains. It was the more homogeneous class and for this reason we used it to illustrate BS structural variability arising from structural homology by choosing three representatives. It should be noted that two of these (chromodomains from mouse HP1 beta, PDB code: 1GUW; and from Polycym, PDB code: 1PDQ) had to be slightly trimmed before identifying them as members of this class.

The cavity insertion class (Fig. 3) included BS from the bromodomain, the tandem Tudor and the WDR5 effectors. For these cases, apart from the tunnel-like cavity, it was hard to identify other common features. Probably the most different case was that of the bromodomain complex as it also displayed features from the narrow-groove class.

Use of the molecular surface representation highlighted a prominent feature in the members of the flat-groove class (Fig. 1): a protruding part from the effector (corresponding to an aromatic ring) anchored the histone peptide, which adopted a hook-like structure at this locus. This motif was also found in the representatives of the narrow-groove class (Fig. 2), although the role of the tryptophan aromatic ring was not so relevant. A more different variant of the motif was also found in the bromodomain member of the cavity-insertion mode (Fig. 3). Interestingly, in all cases the peptide local structure was very similar.

**Automatic structure comparison of the BS.** Use of a hard-sphere representation, which preserves atomic detail, served to evidence a clear structural heterogeneity among BS (Figs. 1–3), between and within the classes previously defined. Both for the flat-groove and narrow-groove representatives we found that the
aromatic cage was relatively well preserved, but the remainder of the BS showed clear intra-class differences (Figs. 1 and 2). When looking at the conserved motif, we found that while the structure of the histone atoms involved looked well preserved, this was not the case for the effector’s contacting atoms (Figs. 1–3).

We decided to use automatic structure comparison methods to assess this structural variability and see whether it was relevant or, on the contrary, it was comparable to experimental noise and therefore negligible.

We did an all-against-all comparison of the BS structures using the program MAMMOTH\(^1\) (see Materials and Methods). For each comparison we obtained an alignment and an associated rmsd (root-mean-square deviation) value. The rmsd was used as a measure of the structural variability between BS. As mentioned before BS from effectors with the same fold, which corresponded to instances of structure divergence, looked relatively similar on visual inspection (Fig. 2). On the contrary, BS from effectors with different folds (e.g., Tudor and PHD effectors from the flat-groove class), which corresponded to examples of structure convergence, looked more different (Fig. 1). Consequently, in these cases similarities could be harder to find by MAMMOTH.

To take into account this heterogeneity in the comparisons we grouped them into two separate classes: INTRADOM and INTERDOM, which corresponded to comparisons between BS from effectors with the same and different folds, respectively. The rmsd distributions for these two classes were compared with (1) the coordinate uncertainty of experimental origin, which varies between 0.1 Å and 1 Å,\(^2\) and (2) the rmsd values resulting from the comparison between different versions of the same effector (apo- and holo-forms, complexes with different ligands), what we called the AUTODOM class.

We found that all INTRADOM values were above 1 Å, therefore the corresponding structural diversity could not be attributed to noise of experimental origin. In accordance with this, we found that INTRADOM and AUTODOM distributions were significantly different (Kolmogorov-Smirnov test: \(p \approx 0\)), although they showed a certain degree of overlap (Fig. 4A). This small overlap was mostly due to the fact that values of AUTODOM above 3 Å corresponded to comparisons involving BS defined from different ligands, more prone to give incorrect alignments. The fact that INTRADOM was different from AUTODOM indicated that structural differences between homologs were larger than those arising from a mere local side-chain rearrangement such as that occurring after substrate binding. To complete the rmsd analysis we computed the fraction of degenerate positions (those positions where the aligned atoms were different, e.g., C and O) present in the alignments (Fig. 4B). We found that AUTODOM and INTRADOM distributions were different (Kolmogorov-Smirnov test: \(p \approx 0\)) although their overlap was larger than in the case of rmsd.

AUTODOM and INTERDOM rmsd distributions were different (Kolmogorov-Smirnov test: \(p \approx 0\)), and only had a small overlap, but INTERDOM and INTRADOM distributions showed a substantial overlap (Fig. 4A) suggesting that the structural variability for convergent and divergent cases was similar. However, inspection of the INTERDOM alignments showed that this was not true. For a set of the most promising cases from this distribution we found that all MAMMOTH alignments were meaningless (when analyzed visually) and their low rmsd values (near 4 Å) were essentially due to the small number of aligned atoms. There was only one exception: the comparison between the human CHD1 tandem chromodomains (PDB code: 2B2W) and the PHD finger from human BPTF (PDB code: 2FSA) (Fig. 5). In this case the alignment superimposed the motif found in the previous section and present in both BS. These results confirmed our visual analysis of the hard-sphere representations showing that structural variability among convergent cases was larger than what was suggested by overall shape similarity. Indeed, while MAMMOTH was able to produce

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Effector domain</th>
<th>Modified residues</th>
<th>PDB codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histone peptides</td>
<td>Chromodomain</td>
<td>H3K4me2K9me2(^+)</td>
<td>1GUW(^+)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H3K9me1</td>
<td>1Q3L</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H3K9me2</td>
<td>1KNA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H3K9me3</td>
<td>1KNE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H3K27me3</td>
<td>1PDQ, 1PF8</td>
</tr>
<tr>
<td>Double</td>
<td>Chromodomain</td>
<td>H3K4me3T3ph</td>
<td>2B2T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H3K4me3R2me2</td>
<td>2B2U</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H3K4me1</td>
<td>2B2V</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H3K4me2</td>
<td>2B2W</td>
</tr>
<tr>
<td>Bromodomain</td>
<td></td>
<td>H4K16ac</td>
<td>1E6l</td>
</tr>
<tr>
<td>PHD finger</td>
<td></td>
<td>H3K4me2</td>
<td>2F5A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H3K4me3</td>
<td>2F6J, 2FU,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2G6Q</td>
</tr>
<tr>
<td>Tandem Tudor</td>
<td></td>
<td>H4K20me2</td>
<td>2I0G</td>
</tr>
<tr>
<td>Double Tudor</td>
<td></td>
<td>H3K4me3</td>
<td>2GFA</td>
</tr>
<tr>
<td>WDR5</td>
<td></td>
<td>H3K4</td>
<td>2C00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H3K4me</td>
<td>2H6K</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H3K4me2</td>
<td>2H6N, 2CNX,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2G99, 2G9A,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2H13</td>
</tr>
<tr>
<td>Non-histone</td>
<td>Chromo Shadow/</td>
<td>Chr. Ass. Fac.</td>
<td>1S4Z</td>
</tr>
<tr>
<td>ligand</td>
<td>Chromodomain(^c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EMPSY</td>
<td>2FMM</td>
</tr>
<tr>
<td>Bromodomain</td>
<td>TAT peptide</td>
<td>p53 C-t</td>
<td>1J4B</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1JSBP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NP1</td>
<td>1WUG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NP2</td>
<td>1WUM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MIB</td>
<td>1Z55A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTR</td>
<td>2D82A</td>
</tr>
</tbody>
</table>

\(^{a}\)Ligand. For histones only the modified, or the normal target, residues are shown. \(^{b}\)PDB code. \(^{c}\)Chromo Shadow domains were treated as chromodomains, as their structural comparison with chromodomains gave excellent alignments.
The resulting multiple structure alignment allowed us to identify two regions in the bound peptides (Fig. 6A): a highly conserved hook-like motif (shown in red) and a non-conserved region, corresponding to the remainder of the peptide (the parts upstream and downstream the conserved motif). Within the non-conserved regions we could distinguish a clear cluster corresponding to peptides binding narrow-groove class members (Fig. 6B). As expected, peptides bound to flat-groove class members showed a larger variability in the non-conserved regions (Fig. 6C). Most of this variability could be attributed to the peptide bound to the tandem chromodomains (PDB code: 2B2T); the similarity between the remaining peptides was closer to that found between narrow-groove members.

The hook-like motif was highly conserved (Table 2), irrespective of the effector’s BS class. It involved the following atoms: the side chain (except the terminal N and its bound methyl or acetyl groups) and the N and Cα main chain atoms from the modified lysine; the main chain atoms of the residue at position -1 relative to the modified lysine; and the C and O main chain atoms of the residue at position -2. An incomplete version of the hook-like motif was also found in the peptides bound to the bromodomain (PDB code: 1E6I) and to the tandem Tudor effector (PDB code: 2IG0), which lacked the residue at position -2 and therefore the corresponding C and O main chain atoms. The motif was not found in the WDR5 complexes, where the peptide adopted a different structure upstream the modified lysine.

Figure 1. Classification of the effector BS. The flat-groove class. In this figure and in the next two figures (Figs. 2 and 3) we show three representative members of each class with their names and PDB codes, plotted using two representations: molecular surface (upper row) and hard sphere (lower row). Shown in yellow are the BS atoms in contact with the histone peptide, and in orange those included in the BS to improve shape representation. The remainder of the effector is represented with a grey mesh. The peptide is represented in green, except for the part corresponding to the conserved hook-like motif, shown in red.

Figure 2. Classification of the effector BS. The narrow-groove class. For an explanation on the molecular representations and color codes, see legend to Figure 1.

reasonable alignments between divergent BS, this was not the case for convergent BS.

Common ligand features. Histone peptides tend to adopt an extended conformation when bound to effectors. Here we explore the degree of conservation of this extended conformation. To this end we superimposed a series of representative histone peptides (PDB codes: 1GUW, 2B2T, 1PDQ, 2FUU, 2G6Q, 2GFA, 1E6I, 2IG0) against the structure of the histone peptide from the histone-chromodomain complex from Drosophila melanogaster (PDB code: 1Q3L), arbitrarily chosen as reference. The resulting multiple structure alignment allowed us to identify two regions in the bound peptides (Fig. 6A): a highly conserved hook-like motif (shown in red) and a non-conserved region, corresponding to the remainder of the peptide (the parts upstream and downstream the conserved motif). Within the non-conserved regions we could distinguish a clear cluster corresponding to peptides binding narrow-groove class members (Fig. 6B). As expected, peptides bound to flat-groove class members showed a larger variability in the non-conserved regions (Fig. 6C). Most of this variability could be attributed to the peptide bound to the tandem chromodomains (PDB code: 2B2T); the similarity between the remaining peptides was closer to that found between narrow-groove members.

The hook-like motif was highly conserved (Table 2), irrespective of the effector’s BS class. It involved the following atoms: the side chain (except the terminal N and its bound methyl or acetyl groups) and the N and Cα main chain atoms from the modified lysine; the main chain atoms of the residue at position -1 relative to the modified lysine; and the C and O main chain atoms of the residue at position -2. An incomplete version of the hook-like motif was also found in the peptides bound to the bromodomain (PDB code: 1E6I) and to the tandem Tudor effector (PDB code: 2IG0), which lacked the residue at position -2 and therefore the corresponding C and O main chain atoms. The motif was not found in the WDR5 complexes, where the peptide adopted a different structure upstream the modified lysine.

Given the high conservation degree of the hook-like motif, we decided to characterize its contribution to the histone-effector interaction. To this end we computed the percentage of accessible surface area (ASA) buried by the motif upon complex formation, relative to the ASA buried by the peptide and by the whole histone. The first percentage was directly obtained from the structures of the complexes. The average ASA buried by the motif was ~113 Å² (Table 2). Relative to the ASA buried by the whole peptide, the values of the hook-like motif are high, between 10% and 25% of the total, indicating that the motif identified also played an important role in the binding affinity of the peptide-effector interaction.

The size of the histone peptides in our dataset varies substantially (Table 1) and may not properly reflect what happens in complexes in which whole histones are involved. For this reason we decided to obtain an estimate of the ASA buried by the motif relative to that buried by the whole histone. The latter could not be directly computed from the structures available and was estimated applying the following assumptions: (i) that the ASA buried by histones had to have an upper threshold, i.e., it had to tend asymptotically to a given value; and (ii) that this value should resemble for all effectors, given the similar sizes of their BS. The first assumption was based on the fact that in our dataset the largest peptides tended to have unbound terminal ends. The second assumption, which implies
an averaging over all binding modes, was motivated by visual inspection of the peptide-effector complexes, which showed that peptide-effector interfaces had similar sizes. A plot of the peptide buried ASA vs. peptide length for the complexes in our dataset showed an asymptotic behavior, with buried ASA approaching 1,160 Å² (Fig. 7) as peptide length increased. We took this value as an approximation of the ASA buried by whole histones upon effector binding. Using this value we found that the contribution of the hook-like motif to an average histone-effector interaction was near 10% (Table 2), confirming its relevance when an approximate but more realistic scenario was considered.

To refine our analysis we broke down the motif’s buried ASA into atomic contributions. The result (Fig. 8) showed a clear general trend defined by the presence of two peaks: one for the C and O main chain atoms of the residue at position -2 relative to the modified lysine, and the other for the lysine side chain itself. However, underlying this trend there was a clear variability that reflected both the BS and the histone sequence variabilities. This was established by grouping the results according to the residue sequence of the hook-like motif (two tripeptides only: ARK and RTK, Fig. 8B). The variability within groups reflected the underlying BS variability: ARK motifs, which were bound to homolog BS (belonging to the narrow-groove class), had smaller variability than RTK motifs, which were bound to BS from the flat-groove class (more structurally heterogeneous). The differences between groups reflected the sequence differences between hook-like motifs.

**Discussion**

The availability of structural information for different histone-effector complexes provides a good opportunity to advance our understanding of the molecular basis of epigenetic regulation, by shedding light on the biophysical/functional properties of this interaction and of its components. In this article we have studied the structural variability of histone-effector complexes focusing on the effector BS, and on the bound histone peptides. We found that there was an incomplete equivalence between our BS-level classes and Taverna and colleagues’ MBS-level classes: their surface-groove class could be found in both our flat-groove (Fig. 1) and narrow-groove (Fig. 2) representatives. This discrepancy is not contradictory, as these classes arise from considering BS at two different levels (whole and part, respectively). It indicates that the BS of effectors can be broken down in two complementary parts, MBS and remainder of the BS, and that for effectors with the same MBS (like chromodomain and PHD) differences in binding specificity/affinity will be determined by the rest of the BS. This idea is supported by both mutagenesis and sequence data. For example, for the second Tudor domain (PDB code: 2GFA) of Jumonji domain-containing protein 2A, mutations Asp945Ala and Asp945Arg reduce and eliminate H3K4me3 binding, respectively. Because this residue is in contact with the histone peptide-but not with the modified lysine, this result is in accordance with the role of specificity-responsibles proposed for BS residues outside the MBS.
Results of the automatic comparison of BS from homolog effectors (BS of the chromodomains from the narrow-groove class) point in the same direction. It would seem that because of the high structural similarity of these BS, which have the same MBS, they would have the same substrate specificity. However, this is not always the case, as shown for the Polycomb and HP1 chromodomains. In spite of both belonging to the same BS class, narrow-groove and MBS class, surface-groove, these chromodomains bind two different substrates, as mentioned before: H3 trimethylated lysines K27 and K9, respectively. Fischle and colleagues have shown that distinct BS features outside the MBS can play an important role determining the different substrate specificities of Polycomb and HP1 chromodomains. In spite of both belonging to the same BS class, narrow-groove and MBS class, surface-groove, these chromodomains bind two different substrates, as mentioned before: H3 trimethylated lysines K27 and K9, respectively. Fischle and colleagues have shown that distinct BS features outside the MBS can play an important role determining the different substrate specificities of Polycomb and HP1 chromodomains. Our results generalize this observation, first by showing that the variability between homolog BS, described by the INTRADOM distribution (Fig. 4A), is larger than experimental noise. Therefore, there are non-trivial structural differences between BS that may result in their having different interaction profiles likely to introduce subtle specificity differences, such as those described in the case of the Polycomb and HP1 chromodomains. In addition, the comparison between the INTRADOM and AUTODOM rmsd

Figure 5. Structure alignment of the BS from the tandem chromodomains of human CHD1 (PDB code: 2B2W) and the PHD finger from human BPTF (PDB code: 2FSA). The alignment was obtained with MAMMOTH, without using the structure of the bound peptides. The BS of the CHD1 chromodomains (represented with a mesh) and the bound peptide are shown in magenta, the remainder of the structure was represented with a green ribbon. The BS of the PHD finger (represented with a continuous molecular surface) and the bound peptide are shown in yellow, the remainder of the structure was represented with a red ribbon. The good coincidence of the peptide structures shows the location of the common structural motif, formed by a protruding tryptophan side chain and a hook-like peptide substructure.
distributions (Fig. 4A) can be used to shed some light on the mechanism underlying the structural differences between homolog BS. It has been shown that distributions equal in nature to the AUTODOM rmsd distribution reflect the native-state dynamics of the protein. The fact that INTRADOM and AUTODOM rmsd distributions are different indicates that structural differences between homolog BS are unlikely to arise from the stabilization, through substrate binding, of common conformational states from their native dynamics. Rather, they are more likely to correspond to states specific to each effector that, in turn, determine the substrate binding mode. In other words, specificity differences between homolog effectors would have a structural component, complemented by the chemical differences resulting from sequence divergence, of which the percentage of non-degenerate sites (Fig. 4B) is an approximate measure. Confirmation of this idea constitutes a challenging problem that would require the use of simulation techniques beyond the structure analysis tools used in this article.

Our analyses of the peptide structures gave a picture consistent with the BS classification: we found a substantial degree of variability arising from differences between the corresponding effectors (Fig. 6B and C). However, we could also identify a structural motif, the hook-like motif, present in most peptides regardless of the effector class (Fig. 6). Analysis of the interaction pattern of this motif allows refining our view on the partition of the BS according to specificity determinants. When the residue sequence of the hook-like motif was RTK we could see a high variability in its interaction pattern (Fig. 8B). As explained before, this variability was related to differences in the binding effectors. This indicates that the structure of atoms nearby, but outside, the MBS are involved in specificity-determining interactions. On the contrary, when the sequence of the hook-like motif was ARK its interaction pattern was well conserved. As all the ARK motif-carrying peptides were bound to homolog effectors (the chromodomains from the narrow-groove class) this confirms that the substrate specificity determinants towards different histone peptides involve effector atoms in contact with histone atoms outside the conserved sequence residue. This is in accordance with the results obtained by Fischle and colleagues in the case of Polycomb and HP1 chromodomains.

Our view on specificity determinants was completed by the analysis of the histone peptide structures and of their local structure propensities. It is known that histone N-terminal tails are intrinsically disordered and undergo a disorder-order transition when binding effectors. Because for interactions involving disordered proteins specificity depends on this transition, we decided to see whether our data could shed some light upon its nature, by focusing on the composition of the disordered state. We restricted our analysis to the disordered state of ARK and RTK tripeptides, as they correspond to the residue sequence of...
K4 and K9 (which includes the ARK motif associated to K9) populates more frequently the helical state, while the first four N-terminal residues (which include the RTK motif associated to K4) populate more frequently the extended structure (Fig. 2). The consistency between (1) Liu and Duan’s simulations and (2) our database statistics strongly supports the idea that the disordered state, and consequently the disorder-order transition, of ARK and RTK are different. The balance between the contribution of this transition and the atomic contacts made upon binding will determine the strength of the histone-effector interaction and the binding specificity of effectors towards given modifications. Experimental results together with results from the aforementioned molecular dynamics simulation suggest that histone modifications could modulate the formation of the histone-effector complex by shifting the histone tail equilibrium population from the helical to the extended state, or vice versa. The resulting effect will combine with the contribution of the interaction between the histone peptides and the binding sites specificity determinants to produce the final binding specificity for the histone-effector interaction.

The hook-like motifs (Table 2) and contribute an average of 44% of the ASA buried by histone peptides upon binding. The high conservation degree of the hook-like motif (Fig. 6 and Table 2) suggested that ARK and RTK tripeptides could have a strong propensity towards this structure. To test whether this was the case we analyzed all the tripeptides with these sequences present in a non-redundant set of the PDB. We found that ARK and RTK adopted a variety of structures (Table 3), many of them different from those found in bound histone peptides. In addition, the structural propensities of these tripeptides varied between them: for example, RTK was most commonly found in the state it adopts in histone-effector complexes than ARK. We also checked the secondary structure of both tripeptides, finding again clear differences: ARK tripeptides were more frequently found as part of α-helices (59%) than β-strands (3%), while for RTK tripeptides the differences between both states where smaller (18 and 11%, respectively). These results are in accordance with recent molecular dynamics simulations of an 18-residue peptide encompassing the first fifteen residues of histone H3. These simulations show that the sequence stretch between K4 and K9 (which includes the ARK motif associated to K9) populates more frequently the helical state, while the first four N-terminal residues (which include the RTK motif associated to K4) populate more frequently the extended structure (Fig. 2 in Liu and Duan). The consistency between (1) Liu and Duan’s simulations and (2) our database statistics strongly supports the idea that the disordered state, and consequently the disorder-order transition, of ARK and RTK are different. The balance between the contribution of this transition and the atomic contacts made upon binding will determine the strength of the histone-effector interaction and the binding specificity of effectors towards given modifications. Experimental results together with results from the aforementioned molecular dynamics simulation suggest that histone modifications could modulate the formation of the histone-effector complex by shifting the histone tail equilibrium population from the helical to the extended state, or vice versa. The resulting effect will combine with the contribution of the interaction between the histone peptides and the binding sites specificity determinants to produce the final binding specificity for the histone-effector interaction.

Figure 7. Histone peptide ASA buried upon effector binding vs. peptide length. The observations corresponding to the different peptides are shown with a triangle. The dashed line represents the curve of equation 1159.5 - (7611.6/(6.0 + peptide length)), obtained after a non-linear fit of the data.
The specificity of whole multidomain epigenetic regulators for given chromatin loci could then be the result of several specific histone-effector interactions, as postulated by Ruthenburg and colleagues in their multivalence model.

We are still far from a quantitative model embracing all these effects and allowing the raising of very general, experimentally testable predictions on the histone-effector interaction. However, our results can already be used to make concrete predictions on how specificity of the histone-effector interaction is modulated. For example, it has been shown that the chromodomains of Polycomb family members display differential binding to H3K27me3. In particular, some of them are able to bind H3K9me3, like the chromodomains from HP1 proteins. For example, Cbx7 chromodomain is able to bind both H3K9me3 and H3K27me3, and Cbx4 chromodomain prefers H3K9me3. Interestingly, in the list of Polycomb BS residues outside the MBS region and their equivalent HP1 residues (first paragraph of the Discussion) we find that Ala28 from Polycomb is paired with Val26 from HP1, a residue shown to play an important role in histone binding. A look at the multiple sequence alignment for the Polycomb family shows that the Cbx4 and Cbx7 residues equivalent to Drosophila Polycomb Ala28 are valines. This strongly suggests that this mutation in one of Polycomb's specificity determinants may play an important role in the specificity shift towards H3K9me3 observed for Cbx4 and Cbx7. While there may be other associated residue changes, this example illustrates how our results can be used to shed light on the regulation of the histone-effector interaction. Also, low-resolution knowledge of the disordered state populated by histone peptides nearby modified residues, such as that provided by our database study, may be used to obtain clues on the nature of the effector BS. For example, peptides with disordered states populating many conformations may require more interactions with the effector BS to form a stable complex than those with less heterogeneous disordered states. This in turn will require larger or deeper BS, able to form more interactions with the histone than flatter BS.

The work presented here focuses on lysine modifications, which represent a subset of all histone tail modifications. A complete reproduction of our analyses for other target residues—Arg,
Lysines, with different modifications. In a very few cases the complexes involved non-histone molecules (proteins and organic molecules) instead of histone peptides. These complexes were included to increase the sampling of the BS variability, although in some cases the automatic alignment procedure could not identify the common parts between them (see below).

The BS atoms are: all effector atoms contacting any ligand atom (atom-atom distance lower than 5 Å), and some neighboring atoms, included to improve the shape representation of the BS.25 The latter are obtained from the cavity pattern of the effector. First, we computed the effector cavities using SURFNET.26 Second, we found all cavities with more than 50% of their atoms in contact with the ligand; all the atoms of these cavities were considered as BS atoms. Finally, atoms from other cavities were added if they were in contact with all the previously included effector atoms (atom-atom distance lower than 5 Å).

The automatic alignment of BS. BS were aligned using the program MAMMOTH.11 This program was conceived for the alignment of Cα-traces and is very fast. Because its alignment algorithm is based on geometrical principles it can be applied to our problem.

To confirm that MAMMOTH alignments were meaningful, we explored manually a large number of them with the package PyMol.27 We found that for the vast majority of cases

Materials and Methods

Structure of the histone-effector complexes. The structures of the complexes used are listed in Table 1, with their PDB codes. The effectors involved were: bromodomain, chromodomain, Tudor, PHD and WDR5. Most of the histone tails were from histone H3 and variants (from different species), although in two cases they were from histone H4. They had residues, mostly lysines, with different modifications. In a very few cases the complexes involved non-histone molecules (proteins and organic molecules) instead of histone peptides. These complexes were included to increase the sampling of the BS variability, although in some cases the automatic alignment procedure could not identify the common parts between them (see below).

The BS. BS atoms are: all effector atoms contacting any ligand atom (atom-atom distance lower than 5 Å), and some neighboring atoms, included to improve the shape representation of the BS.25 The latter are obtained from the cavity pattern of the effector. First, we computed the effector cavities using SURFNET.26 Second, we found all cavities with more than 50% of their atoms in contact with the ligand; all the atoms of these cavities were considered as BS atoms. Finally, atoms from other cavities were added if they were in contact with all the previously included effector atoms (atom-atom distance lower than 5 Å).

To confirm that MAMMOTH alignments were meaningful, we explored manually a large number of them with the package PyMol.27 We found that for the vast majority of cases
MAMMOTH would give reasonable results, aligning sets of atoms that defined a similar shape. As an example, in Figure 9 we show the alignment between two chromodomains binding two slightly different histone peptides.

It has to be mentioned that the contouring atoms that define a BS may remain constant or vary among different versions of a given structure. That is, at some locations we may find different atoms, e.g., an aromatic carbon or a polar nitrogen, depending on the BS version considered. This degeneracy has three different origins: sequence divergence, natural dynamics of the protein and technical indeterminacies. When comparing BS, degenerate positions may be excluded or included in the alignment. We decided to include them (i.e., MAMMOTH was allowed to align any pair of atoms, regardless of their nature) and complement our results with a measure of their abundance, i.e., the percentage of non-degenerate positions in the alignment. This may help to understand the degree of structural and chemical divergence at the BS.

Peptide alignment. Eight representative peptides were aligned following a semi-automatic procedure. We aligned eight peptides (taken from the complexes with PDB codes: 1GUW, 2B2T, 1PDQ, 2FUU, 2G6Q, 2GFA, 1E6I, 2IG0) against the histone peptide in the histone-chromodomain complex from Drosophila melanogaster (PDB code: 1Q3L), which was arbitrarily chosen. For each comparison the protocol followed was: (1) visually identify an initial set of equivalent atoms in both peptides; (2) superimpose this atom set using the Kabsch algorithm and if the rmsd is above 1.1 Å the alignment is discarded; (3) explore visually the resulting alignment and identify any possible additional atom pairs; (4) if there are new possible pairs, add them to the original atom set and go to step (2), otherwise the protocol is finished.

Atomic surface area computations. Buried atomic accessible surface area (ASA) is an atomic contact descriptor related to the contact’s free energy contribution and for this reason constitutes a valuable tool in the study of molecular interactions. The percentage of accessible surface area buried by the hook-like motif upon complex formation was computed as follows: 

\[
\text{ASA}_{\text{hook}}(\text{peptide in isolation}) - \text{ASA}_{\text{hook}}(\text{complexed peptide}) / \text{ASA}_{\text{histone}}(\text{complexed peptide}),
\]

where \(\text{ASA}_{\text{histone}}\) was the estimate of the ASA buried by the whole histone upon complex formation (it was obtained as explained in the RESULTS section). The two peptide ASA values, \(\text{ASA}_{\text{hook}}(\text{peptide in isolation})\) and \(\text{ASA}_{\text{hook}}(\text{complexed peptide})\), were obtained with the program NACCESS.

Structural propensity of the ARK and RTK tripeptides. In the hook-like motif we can distinguish two sources of structural variety: the lysine side chain torsional angles, and the main chain angles \(\phi_i, \psi_i\) and \(\phi_i\). We focused our analysis on the latter, which are related to the main structural features of the disorder-order transition experienced by histone tails upon binding. To see if their highly conserved states on bound peptides were due to intrinsic sequence propensities, we looked in a non-redundant subset of the PDB structural database for instances of ARK and RTK tripeptides (corresponding to the hook-like motifs in our dataset) and obtained the values of their \(\phi_i, \psi_i\) and \(\phi_i\) torsionals \((60 < \phi_i < 180; -60 < t < 60; -180 < t < -60)\) from the output of the DSSP program.

Conclusions

Using structural analyses of a series of histone-effector complexes we have characterized the structural variability of the two main components of the histone-effector interaction: histone tails and the binding site of effector modules. The results of this analysis have crystallized in a coherent classification for the latter that allows us to propose that, in general, the BS of effector domains is partitioned in two different specificity determinants (the MBS and the remainder of the BS). In addition, structural analysis of the bound histone peptides and their sequences led to the identification of an additional specificity determinant: the disorder-order transition. This transition, which takes place upon histone binding, would be a specific property of the sequence nearby the post-translationally modified histone residue (e.g., it would be different for H3K4 and H3K9). Overall, our results contribute to clarify the specificity origins for the histone-effector interaction.

Acknowledgements

We thank R. Jackson and the members of the Molecular Modeling and Bioinformatics group for helpful comments. We also thank both referees for constructive comments on our work. This work was supported by the Spanish Ministerio de Educación y Ciencia (grant numbers BIO2006-15557, BFU2006-01493, CSD2006-00049); and the Spanish Ministerio de Ciencia e Innovación (grant numbers BFU2009-11527, BFU2009-11144).

References


