Research Article

Structural basis of molecular recognition of helical histone H3 tail by PHD finger domains

Alessio Bortoluzzi*, Anastasia Amato*, Xavier Lucas, Manuel Blank and Alessio Ciulli

Division of Biological Chemistry and Drug Discovery, School of Life Sciences, University of Dundee, James Black Centre, Dow Street, Dundee DD1 5EH, U.K.

Correspondence: Alessio Ciulli (a.ciulli@dundee.ac.uk)

The plant homeodomain (PHD) fingers are among the largest family of epigenetic domains, first characterized as readers of methylated H3K4. Readout of histone post-translational modifications by PHDs has been the subject of intense investigation; however, less is known about the recognition of secondary structure features within the histone tail itself. We solved the crystal structure of the PHD finger of the bromodomain adjacent to zinc finger 2A [BAZ2A, also known as TIP5 (TTF-I/interacting protein 5)] in complex with unmodified N-terminal histone H3 tail. The peptide is bound in a helical folded-back conformation after K4, induced by an acidic patch on the protein surface that prevents peptide binding in an extended conformation. Structural bioinformatics analyses identify a conserved Asp/Glu residue that we name ‘acidic wall’, found to be mutually exclusive with the conserved Trp for K4Me recognition. Neutralization or inversion of the charges at the acidic wall patch in BAZ2A, and homologous BAZ2B, weakened H3 binding. We identify simple mutations on H3 that strikingly enhance or reduce binding, as a result of their stabilization or destabilization of H3 helicity. Our work unravels the structural basis for binding of the helical H3 tail by PHD fingers and suggests that molecular recognition of secondary structure motifs within histone tails could represent an additional layer of regulation in epigenetic processes.

Introduction

The plant homeodomain (PHD) finger is one of the largest families of epigenetic reader domains present in chromatin-related proteins, with over 170 PHD fingers identified in the human genome. Early pioneering studies led to PHD fingers being classified as domains that specifically recognize histone H3 trimethylated at K4. However, the diversity of PHD fingers in terms of their ability to recognize a wide array of post-translational modifications (PTMs) and unmodified tails has now become apparent. Several PHDs have been characterized that recognize different PTMs on the H3 tail, including di- and tri-methylation of K4, trimethylation of K9, acetylation of K14 and trimethylation of K36 as well as PTMs on H4 such as acetylation. An additional layer of complexity in the molecular recognition by PHD fingers is imparted by the recurrent presence of adjacent domains that aid combinatorial, multivalent readout of histone tails, intra- or inter-nucleosomal. Indeed, PHD fingers are often found in close proximity with a bromodomain (BRD) in complex proximity with a bromodomain (BRD), as well as other PHD fingers, bromo-adjacent homology domains, tudor domains and chromodomains. Structural studies have elucidated diverse modes of combinatorial readout by PHD fingers and their tandem domains for individual and multiple histone tails, which typically involve recognition of the peptide in a fully extended conformation. While PTM-specific and combinatorial readout modalities of histone tails are well understood, much less is known about the recognition of secondary structure features within the histone tail itself.

Members of the BAZ family of proteins, which includes BAZ1A, also known as Acf1, BAZ1B, also known as Wstf, BAZ2A, also known as TIP5 (TTF-I/interacting protein 5), BAZ2B, are all characterized by the presence of a PHD–BRD tandem module at...
their C-terminus. BAZ2A is the best characterized member of the BAZ family from a functional standpoint. BAZ2A binds to the ATPase SNF2h (sucrose nonfermenting protein 2 homolog) to form the chromatin remodeling complex NoRC (nucleolar remodeling complex), which plays an essential role in silencing ribosomal DNA (rDNA) genes [27]. Experiments performed with truncated versions of BAZ2A showed that the PHD–BRD module plays an important role in NoRC formation, with the PHD domain being required for interaction with the nucleosome to trigger transcriptional silencing of rDNA [28]. In a recent study, BAZ2A was found to be overexpressed in prostate cancer and a role was proposed for the protein in establishing epigenetic alterations that favor an aggressive phenotype of the cancer [29]. The related protein BAZ2B [30] is yet poorly characterized and its biological role remains unclear. We recently biochemically and structurally characterized the PHD fingers and BRDs of both BAZ2A and BAZ2B, and identified the N-terminal tail of histone H3 as the preferred binding partner of the PHD domains [26]. Structural studies with PHD–BRD tandem constructs have pointed to rather elongated and rigid structures with the two domains probably recognizing distinct regions of H3 histone tails independently [26]. NMR spectroscopy has been combined with computational studies to throw light on the molecular recognition features of histone H3K14ac recognition by the BAZ2B BRD [31]. However, the complete molecular picture of H3 tail recognition by the PHD fingers of BAZ2A and BAZ2B had remained elusive.

**Results**

**BAZ2A PHD recognizes H3 tails in a helical fold**

To elucidate the molecular detail of histone H3 N-terminal tail recognition, we solved the crystal structure of ARTKQTARKS (H3 10-mer) bound to BAZ2A PHD (Figure 1A–C; see Table 1 for X-ray data collection and refinement statistics). The peptide residues A1–K4 form an antiparallel β-sheet with the first β-strand of BAZ2A PHD, anchored by backbone hydrogen bonds with residues D1688, L1692, L1693, P1714 and G1716 (Figure 1C). This region of the peptide is found essentially in the same conformation observed in the crystal structure of BAZ2A PHD with bound H3 5-mer (ARTKQ) [26]. The methyl groups of A1 and T3 contribute hydrophobic interactions to peptide binding, and further contributions are brought by the hydrogen bonds and electrostatic interactions of R2 and K4 side chains (Figure 1C). However, starting from K4, the peptide adopts a helical fold that extends at least until R8, forming a complete loop of an α-helix (Figure 1A). The canonical intrapeptide i to i + 4 backbone hydrogen bonds stabilize the helix loop i.e. T3 to A7 and K4 to R8 (Figure 1C). Two additional side chain-to-backbone intramolecular hydrogen bonds are formed, one between the T3 hydroxyl group and the amino group of T6, and a second one between the hydroxyl group of T6 and the amino group of R2 (Figure 1C). Phosphorylation of T3 and methylation of R2 had been shown to lower the binding between BAZ2A PHD and H3 peptide [26], consistent with disruption of these interactions. The electron density for the side chain of R8 is incomplete after Cβ (Figure 1B). There is no interpretable density for K9 and S10, suggesting that these are disordered (Figure 1B). The fold assumed by the peptide is not influenced by crystal contacts. Inspection of the binding pockets in each of the four chains of the asymmetric unit reveals that the histone-binding sites of chains A and D are both occupied by H3 10-mer and are free from crystal contacts that might interfere with or modulate the secondary structure of the peptide itself. Conversely, crystal packing occludes the binding sites of chains B and C and no peptide is found bound to these protomers.

BAZ2A PHD, and its homologous BAZ2B PHD, each binds H3 10-mer with an affinity ~4-fold higher compared with H3 5-mer (Figure 1D, Supplementary Figure S1 and Table 2). However strikingly, the structure shows no direct interactions between residues T6–S10 and the protein, besides a potential long-range hydrophobic contact between the A7 methyl group and the L1693 side chain. We thus hypothesized that the extra affinity observed with the longer peptide could arise from intramolecular stabilization of its helical fold that helps to avoid clashes with the protein. Indeed, the structure of BAZ2A PHD would be incompatible with a fully extended binding mode of H3 that is commonly observed in PHD-bound crystal structures (Supplementary Figure S3). The 3₁₀ helix of BAZ2A PHD blocks H3 from binding in such an extended conformation, forcing it to fold back (Supplementary Figure S3). Consistent with these observations, shorter tetrameric peptides ARTK and ARTA, designed to reduce steric clashes with the 3₁₀ helix, bound tighter than H3 5-mer to the BAZ2A PHD domain, and remarkably ARTA bound with comparable affinity to H3 10-mer (Figure 1D, Supplementary Figure S1 and Table 2).
Our structural and biophysical data point to an important role of the H3 tail helicity in the recognition of PHD fingers. To assess the prevalence of this recognition mode, we inspected all structures of PHD fingers in complex with H3 peptides deposited in the Protein Data Bank (PDB). Our analysis revealed that the conformation assumed by residues A1-K4 of H3 upon binding to a PHD finger is normally extended and relatively well conserved. In contrast, the folding of the peptide from K4 onwards varies from a completely extended conformation, e.g. H3 N-terminal peptide bound to the PHD domain of ING2 (PDB: 2G6Q [5]), to an α-helix, e.g. H3 N-terminal peptide bound to the double PHD finger (DPF) of MOZ (PDB: 4LK9 [15]). We identified three possible conformations that an H3 N-terminal peptide can adopt when bound to a PHD finger: helical, bent and fully extended (Figure 2A and Supplementary Figure S4). Interestingly, H3 assumes a helical fold when in complex with PHD fingers that harbor a short helical turn or loop just before the first β-strand. This is a 3_{10} helix in the case of BAZ2A PHD (Figure 1). We noted that the 3_{10} helix is particularly acidic in BAZ2A PHD and in its close homolog BAZ2B PHD, comprising D1688 and E1689 in BAZ2A and E1943 and E1944 in BAZ2B. To investigate the conservation of this structural feature, we performed a multiple sequence alignment with PHD fingers whose structure was solved in complex with an H3 N-terminal peptide (Figure 2A and see Supplementary Figure S4A for full alignment). We observed that the PHD fingers of BAZ2A, UHRF1, MOZ and DPF3 all have a conserved acidic residue in the position corresponding to E1689 of BAZ2A PHD, and all recognize H3 in a folded-back helical conformation starting from K4 onwards (Figure 2B). This acidic residue, which we name ‘acidic wall’, is also structurally conserved, as in all cases it is positioned against the bottom of the first loop formed by H3 (Figure 2B). Topological conservation suggests that the negatively charged carboxylate may help to stabilize the positive dipole of the N-terminus of the helix [32]. In contrast with the

Figure 1. Structural basis of H3 recognition by BAZ2A PHD. (A–C) Crystal structure of BAZ2A PHD (shown in gray) in complex with H3 10-mer (shown in green). (A) Surface and ribbon representation of BAZ2A PHD (regions of positive and negative electrostatic potential are shown in blue and red, respectively) in complex with H3 10-mer shown in a ribbon and stick representation. Residues of the H3 10-mer peptide are labeled. (B) The 2Fo–Fc map contoured at 1σ (shown in gray) for H3 10-mer. (C) Close-up view of the interaction between the BAZ2A PHD and the H3 10-mer peptide. Residues of BAZ2A PHD interacting with the H3 10-mer peptide are shown in a stick representation and labeled in black. Residues of the H3 10-mer peptide are labeled in red. (D) ITC-binding curves of different H3-derived peptides titrated into BAZ2A PHD.
recognition of the helical H3 tail, the bent conformation of H3 bound to the PHD appears to be stabilized by a different set of interactions (Supplementary Figure S4C).

**Prevalence of helical H3 N-terminal tail-recognizing human PHD fingers**

It is remarkable that PHD fingers recognizing H3 in a bent or extended conformation do not have an acidic residue in the position corresponding to E1689 of BAZ2A PHD (Figure 2A and Supplementary Figure S4). To investigate the prevalence of the acidic wall residues in all human PHD fingers, we extended our bioinformatics analysis to the entire human genome [1]. We found that 36 of the 172 sequences annotated as PHD fingers have an acidic residue in the position that corresponds to E1689 of BAZ2A (Supplementary Figure S5). Among these are all the PHD fingers of the BAZ family, CREBBP [33] and the homologous EP300 [34], all members of the DPF family of proteins: DPF1, DPF2 and DPF3, as well as members of the KDM5/JARID1 histone

---

**Table 1 Crystallographic data processing and refinement statistics**

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>BAZ2A PHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ligand</td>
<td>ARTKQTARKS</td>
</tr>
<tr>
<td>Beamline</td>
<td>ID29</td>
</tr>
<tr>
<td>Synchrotron</td>
<td>ESRF</td>
</tr>
</tbody>
</table>

**Processing statistics**

| Space group     | P4_2_2_2 |
| a, b, c (Å)     | 72.61, 72.61, 99.43 |
| α, β, γ (°)     | 90.0, 90.0, 90.0 |
| Resolution limits (Å) | 45.62–2.4 (2.49–2.4) |
| Unique observations | 10,901 (1132) |
| Completeness (%) | 99.6 (100) |
| Redundancy      | 5.4 (5.8) |
| Rmerge (%)      | 10.2 (72.8) |
| R (i/σ)         | 9.3 (2.3) |
| CC1/2 (%)       | 99.4 (71.8) |
| Wavelength      | 0.9762   |

**Refinement statistics**

| Resolution limits (Å) | 58.64–2.4 (2.462–2.4) |
| Rwork/Rfree (%)       | 18.5 (25.0)/23.5 (29.6) |
| Number of reflections | 10,309 (769) |
| Number of atoms       | 1860 |
| Protein/other/solvent | 1740/24/96 |
| Average B factors (Å²) | 52.13 |
| RMSD bond (Å)         | 0.01 |
| RMSD angle (°)        | 1.50 |
| Ramachandran statistics |          |
| Favored (%)           | 97.2 |
| Allowed (%)           | 2.8 |
| Outliers (%)          | 0.0 |
| PDB code              | 5T8R    |
lysine demethylase family: KDM5A, KDM5B, KDM5C and KDM5D [35,36]. Interestingly, we noted that, in all four KDM5 members, only the first PHD domain, which like BAZ2A/B recognizes unmodified K4, but not the second or third, has an acidic residue at this position, and this is mutually exclusive with the presence of the conserved tryptophan residue characteristic of the aromatic cage for methyl-K4 recognition (Figure 2C) [37]. Indeed, only 5 of the 36 sequences containing the acidic wall residue also contain this tryptophan (Supplementary Figure S5). Based on this observation, we postulate that there could be a level of incompatibility between methyl-lysine readout and helical H3 recognition by PHD finger domains. Five PHD fingers bear both an acidic wall residue and the tryptophan needed for methyl-K4 recognition: ASH2L [38], the MLL2 and MLL3 members of the KMT2 family of lysine methyltransferases [39], PHF20 [40] and UBR7 (Supplementary Figure S5). Structural information is available only for ASH2L PHD, which unveils an atypical PHD fold with only one zinc ion coordinated, suggesting that the ASH2L PHD structure is incompatible with histone binding [38]. The remaining four PHD fingers are poorly characterized, their substrate specificity is not known and it is difficult to conclude if they represent genuine exceptions to the observed mutual exclusivity between acidic wall residue and conserved Trp residue.

Characterization of the interaction between BAZ2A and BAZ2B PHD with H3 N-terminal tail by NMR

The small PHDs (~6.5 kDa) of both BAZ2A and BAZ2B yielded high-quality [15N-1H] heteronuclear single-quantum coherence (HSQC) spectra (Figure 3 and Supplementary Figure S6), and all the backbone amide
protons were readily assigned in both constructs except for the first serine residue of BAZ2A PHD [Biological Magnetic Resonance Bank (BMRB) deposition nos 26754 and 25988, for BAZ2A PHD and BAZ2B PHD, respectively]. First, we validated our co-crystal structure of BAZ2A PHD in complex with H3 10-mer. [15N-1H]-HSQC spectra were recorded on a sample of 15N-labeled BAZ2A PHD at increasing concentrations of H3 10-mer (Figure 3). The shifts observed were quantified and mapped on to the structure of the BAZ2A PHD-H3 10-mer complex (Figure 4A). Strong and moderate shifts were found to cluster at β1 and at the 310 helix, with the acidic patch residues D1688 and E1689 giving some of the strongest shifts (Figures 3 and 4A). Additional shifts that recapitulate the contacts observed in the crystal structure include G1716, whose carbonyl group engages in a hydrogen bond with the H3 N-terminus and at the protein N-terminus close to the K4 side

Figure 2. ‘Acidic wall’ residue is conserved among PHD fingers that recognize helical H3 tail. (A) Sequence alignment of PHD fingers whose structure was solved in complex with an H3 N-terminal tail peptide. The column corresponding to E1689 of BAZ2A is highlighted through the alignment with a red box, and Asp or Glu residues in this column are colored in red. The column corresponding to the absolutely conserved tryptophan of PHD fingers that recognize methylated-K4 is highlighted through the alignment with a black box, and tryptophan residues in this column are colored in magenta. PHD fingers that induce the H3 tail to adopt a helical (cyan box), bent (green box) or extended (magenta box) fold are grouped (see Supplementary Figure S4 for full alignment). (B) Structures of PHD fingers (gray cartoon) that have an acidic residue (shown in red) in the position corresponding to E1689 of BAZ2A and that induce the H3 N-terminal peptide (green cartoon) to adopt a helical folded-back conformation. (C) Sequence alignment of PHD fingers of KDM5 proteins. The columns corresponding to E1689 of BAZ2A and to the absolutely conserved tryptophan of PHD fingers that recognize methylated-K4 are highlighted through the alignment as described in A (see Supplementary Figure S5 for full alignment).
chain (Figure 4A). Other shifts are observed for residues of BAZ2A PHD relatively far from the H3 10-mer binding site, e.g. G1696, on a loop that links β1 and β2 strands, and I1703, on a short helix after β2 (Figure 4A). Next, we applied the same procedure to yield a chemical shift perturbation (CSP) histogram and corresponding heat map representative of the binding site of the shorter H3 5-mer (Figure 4B). The chemical shift changes induced by H3 5-mer and H3 10-mer closely overlap, showing minor differences only at the N-terminus of BAZ2A PHD where the H3 10-mer induces additional shifts compared with H3 5-mer (Figure 4A,B). Importantly, we did not observe extra shift clusters that could suggest the presence of additional binding sites exploited by the longer H3 10-mer, consistent with the binding mode observed in our crystal structure. Finally, we studied the binding of the H3 N-terminal tail to BAZ2B PHD by NMR. As in BAZ2A PHD, BAZ2B PHD also harbors an acidic wall residue, E1944, and a 310 helix positioned just before β1 (Figure 4). We found that the chemical shift changes induced by H3 5-mer on BAZ2B PHD (Figure 4C and Supplementary Figure S6B) are remarkably consistent with the ones observed for BAZ2A PHD and most shifts map at equivalent positions in the two PHD fingers (Figure 4), including the acidic wall. Overall, the NMR data suggest a probably conserved molecular recognition of the H3 N-terminal tail by the homologous BAZ2A/B PHD fingers (sequence identity of 66%).

**Role of the acidic wall residue of BAZ2A and BAZ2B PHD fingers in H3 N-terminal tail recognition**

To investigate the role of the acidic patch in H3 N-terminal tail recognition, we mutated E1689 of BAZ2A PHD to Gln and Lys, aiming to neutralize and invert, respectively, the negative charge of the acidic wall side chain. Equivalent mutations were also introduced at the acidic wall of BAZ2B PHD, namely E1944Q and E1944K. Correct folding of the resulting mutants was confirmed by 1H 1D NMR spectra (Supplementary Figure S7). Mutant proteins were compared with wild type (WT) for their ability to bind H3 10-mer peptide by isothermal titration calorimetry (ITC; Figure 5, Supplementary Figure S2 and Table 3). Mutation of the acidic wall to Gln led to a decrease in binding affinity, up to 8-fold with BAZ2B PHD. The effect was even more pronounced when the charge was inverted, as the E1944K mutation completely abrogated binding (Table 3 and Figure 5B). Mutation of the acidic wall residue in BAZ2A PHD also affected the thermodynamic parameters of H3 binding, albeit more moderately than for BAZ2B (Figure 5). Specifically, the E1689Q mutation weakened the binding affinity by ~2-fold, whereas the E1689K mutant showed a loss of binding affinity of
Figure 4. Characterization of the interaction between H3 N-terminal tail and BAZ2A/B PHD in solution by NMR. Chemical shift differences induced by H3-derived peptides on BAZ2A/B PHDs were weighted as described in the Experimental section and plotted against BAZ2A/B PHD sequences. The resulting histograms were used to group residues based on the extent of their CSPs: weak (weighted chemical shift difference value equal or above the average chemical shift), medium (equal or above the average chemical shift plus the standard deviation) and strong (equal or above the average chemical shift plus two times the standard deviation). The CSPs observed were mapped on BAZ2A/B PHDs structures (PDB: 5T8R and 4QF3, respectively) by coloring residues with weak shifts in yellow, medium in orange and strong in red. Residues with a weighted chemical shift difference value lower than the average chemical shift are in white. The H3 10-mer peptide is shown as sticks and colored in green and its residues are labeled in red. In the middle panel, the peptide is omitted for clarity.
2.4-fold (Table 3 and Figure 5A). We noted that in BAZ2A PHD, the residue just preceding E1689 is also acidic (D1688), and in our crystal structure, its side chain forms one side of the pocket that accommodates the K4 side chain of the H3 peptide (Figure 1C). Superposition with other PHD structures bound to the helical H3 tail suggested that the R8 side chain of H3 points backward toward the acidic patch of BAZ2A, and could form a salt bridge with the carboxylate group of the D1688 side chain (Figure 1C). Moreover, in the NMR HSQC spectra, the amide NH of D1688 exhibited large chemical shifts in the presence of the H3 peptide (Figure 3). Based on these observations, we hypothesized that this residue could also be important for binding and could potentially compensate the E1689Q mutation. We therefore designed and expressed a double mutant D1688N/E1689Q to fully neutralize the negative charges on the 310 helix of the acidic wall. This double mutation drastically affected the binding of BAZ2A PHD toward the cognate H3 histone peptide, reducing the binding affinity by 17-fold (Figure 5A and Supplementary Figure S2). Taken together, these data demonstrate that the

Table 3 Summary of the thermodynamic-binding parameters for complex formation between H3 WT 10-mer peptide and mutant BAZ2A/B PHD

<table>
<thead>
<tr>
<th>Protein</th>
<th>$K_D$ ($\mu$M)</th>
<th>$N$</th>
<th>$\Delta H$ (kcal/mol)</th>
<th>$T\Delta S$ (kcal/mol)</th>
<th>$\Delta G$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAZ2A PHD wt</td>
<td>48 ± 2</td>
<td>1.29 ± 0.01</td>
<td>$-7.9 \pm 0.1$</td>
<td>$-2.0 \pm 0.1$</td>
<td>$-5.90 \pm 0.02$</td>
</tr>
<tr>
<td>BAZ2A PHD E1689Q</td>
<td>83 ± 7</td>
<td>1.02 ± 0.03</td>
<td>$-9.6 \pm 0.7$</td>
<td>$-4.0 \pm 0.8$</td>
<td>$-5.57 \pm 0.05$</td>
</tr>
<tr>
<td>BAZ2A PHD E1689K*</td>
<td>116 ± 6</td>
<td>1.1 ± 0.4</td>
<td>$-7.6 \pm 2.5$</td>
<td>$-2.2 \pm 0.4$</td>
<td>$-5.37 \pm 0.03$</td>
</tr>
<tr>
<td>BAZ2A PHD D1688N/E1689Q</td>
<td>810 ± 50</td>
<td>1$^*$</td>
<td>$-9.0 \pm 1.2$</td>
<td>$-4.8 \pm 1.2$</td>
<td>$-4.22 \pm 0.04$</td>
</tr>
<tr>
<td>BAZ2B PHD wt</td>
<td>47 ± 2</td>
<td>1.17 ± 0.02</td>
<td>$-8.3 \pm 0.2$</td>
<td>$-2.4 \pm 0.2$</td>
<td>$-5.90 \pm 0.03$</td>
</tr>
<tr>
<td>BAZ2B PHD E1944Q</td>
<td>660 ± 60</td>
<td>1$^*$</td>
<td>$-8.5 \pm 1.1$</td>
<td>$-4.1 \pm 1.1$</td>
<td>$-4.34 \pm 0.06$</td>
</tr>
<tr>
<td>BAZ2B PHD E1944K*</td>
<td>&gt;1000</td>
<td>N.D.</td>
<td>N.D.</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
</tbody>
</table>

*Titration performed in duplicate.

$^*$N was fixed to 1 during the data fitting.
negatively charged patch corresponding to the acidic wall is an important feature for the recognition of the H3 N-terminal tail by the PHD fingers of BAZ2A and BAZ2B.

Changes in H3 N-terminal tail helicity correlate with different binding affinities for BAZ2A and BAZ2B PHD fingers

To gain a better understanding of the histone molecular recognition, we investigated the energetic contribution of different H3 residues in binding to the PHD fingers of BAZ2A and BAZ2B. We performed an alanine scan where residues 2–6 of the H3 10-mer were mutated individually to alanine and the resulting mutant peptides tested for binding with BAZ2A PHD and BAZ2B PHD by ITC (Supplementary Figure S1 and Table 2). The R2A and T3A mutations abolished binding. The K4A mutation did not affect the binding affinity with BAZ2A PHD and even increased the affinity toward BAZ2B PHD. The Q5A mutation improved binding, and the simultaneous introduction of K4A and Q5A mutations remarkably increased binding affinities by 4-fold (BAZ2A) and 15-fold (BAZ2B) (Table 2). Finally, T6A did not affect the binding affinity of H3 10-mer toward either protein. Our data show that K4-T6 residues are not critical for binding to the PHD fingers of BAZ2A and BAZ2B, while R2-T3 are crucial. These results are consistent with those recently reported by Chakravarty et al. [37] for BAZ2A PHD and the first PHD domain of KDM5B but are distinct from the results of the first PHD of AIRE, which is known to bind H3 in an extended conformation. In that case, the T3A mutation was tolerated, whereas the K4A mutation abolished binding [37].

Indeed, alanine has the highest helix propensity among natural amino acids [41]. To test this hypothesis, the role of the K4A and Q5A mutations in the helical propensity and stability of H3 10-mer was studied by molecular dynamics (MD) simulations (Figure 6). First, we modeled H3 10-mer in the context of the complex with BAZ2A PHD (Figure 6A, left panel, and B,C). The helical character of each amino acid during the last 60 ns of simulation, reported as a percentage of time with secondary structure of 310- or π-helix, is shown in Figure 6A (in blue). Residues K4-T6 are stabilized ~25% of the time as a helical structure. The tendency decreases rapidly after T6. A superposition of the last frame of each replica of the simulation shows that, upon unfolding, the C-terminus of H3 10-mer is naturally flexible and disordered (Figure 6B), in agreement with the lack of electron density observed at residues 9 and 10 in the crystal structure (Figure 1B).

An analysis of the intramolecular hydrogen-bond contacts occurring within the peptide during the simulation shows that the T3–T6 contact observed in the crystal structure is persistent and well conserved (Supplementary Figure S9). Remarkably, introducing alanine residues at positions 4 and 5 to generate the H3 10-mer AA peptide induces a significant stabilization of the helix along the simulation (P < 0.002), which is present over 60% of the time for residues K4-A7 and still over 25% beyond and up to K9 (Figure 6A, left panel, red). The intramolecular hydrogen-bond contacts are consequently strengthened during the simulation and involve residues beyond T6 forming a clear ‘i to i + 4’ pattern characteristic of the α-helix (Supplementary Figure S9).

We hypothesized that the observed increase in helical stability upon alanine mutation could also be reflected in their unbound state. To analyze this effect, we modeled both peptides in aqueous solution (Figure 6A, right panel, and D,E). In the absence of the PHD protein, there is still some helicity, albeit weak, persisting in the WT peptide (~5% of the time). The helical character of the peptide in the unbound state was consistently increased by the introduction of K4A and Q5A mutations to 12% of the time (Figure 6A, right panel). To
Further investigate the relationship between the helical propensity of H3 10-mer and binding affinity toward BAZ2A/B PHDs, we aimed to reduce the peptide helicity by replacing K4 and Q5 with a Gly residue. Indeed, excluding proline, glycine has the lowest helix propensity among natural amino acids [41]. The resulting H3 10-mer GG mutant peptide (ARTGGTARKS) showed markedly reduced binding affinity toward both BAZ2A/B PHD fingers (Supplementary Figure S1 and Table 2). ITC data revealed that the decreases in binding affinity of the H3 10-mer GG mutant are contributed entirely by large entropic penalties (Table 2), consistent with a significant reduction in conformational freedom of the peptide upon binding relative to H3 10-mer WT or H3 10-mer AA mutant peptides. MD simulations further showed a significant weakening of the intramolecular hydrogen-bond network along with a modest decrease in the helical character in the H3 10-mer GG peptide compared with H3 10-mer WT and AA mutant (Figure 6A and Supplementary Figures S9 and S10). Taken together, our data reveals that H3 tail helicity plays an important role in recognition by the PHD domains of BAZ2A/B.

Circular dichroism confirms helicity of H3 N-terminal tail in solution

The helical content of H3 10-mer WT, H3 10-mer AA and H3 10-mer GG peptides in solution was investigated by circular dichroism (CD). The CD spectra recorded in water displayed a depth between 195 and 200 nm (Supplementary Figure S11, top panel), which is a characteristic of disordered proteins or peptides [42]. This is in agreement with the MD results where all three peptides rapidly lost the helical fold used as starting conformation and showed only a modest helical content along the simulation (Figure 6A, right panel). To investigate the propensity of the three peptides to adopt an α-helix fold, CD spectra were recorded at increasing concentrations of 2,2,2-trifluoroethanol (TFE; Supplementary Figure S11). TFE is known to stabilize the helical fold of peptides and is often used to assess their helical propensity [43–46]. The CD spectra of the three peptides obtained at different TFE concentrations were deconvoluted (Supplementary Tables S1–S3) and
the α-helix content found was plotted against TFE concentration (Figure 7). During the TFE titration, the α-helix content increases from 0% to ∼10% for all three peptides but with different trends. Indeed, between 40 and 60% TFE, the H3 10-mer AA peptide has the highest α-helix content, followed by H3 10-mer WT and then H3 10-mer GG. The trends observed are in agreement with the expectation that the K4A and Q5A mutations would increase helical propensity of the H3 10-mer peptide and K4G and Q5G reduce it.

Discussion

Epigenetic regulatory processes modulate human physiology and disease; thus, reaching a comprehensive understanding of their molecular basis is important. Molecular recognition of secondary structural features within histone tails by epigenetic reader domains has received little attention to date. Herein, we have examined the structural and biophysical basis for the recognition of the helical histone H3 tail by PHD fingers, using the PHDs of BAZ2A and BAZ2B as the model system.

Our structural insights into the molecular recognition of histone H3 by the BAZ2A and BAZ2B PHD fingers add to the emerging evidence for specific recognition of the helical H3 tail by this reader family, provided by peptide-bound structures recently solved for the PHD fingers of UHRF1, MOZ and DPF3 (Figure 2B). Identification of a strict conservation for an Asp/Glu residue at the acidic wall position on these family members suggests a simple consensus signature for this subclass of PHD domains. Sequence alignment of the whole human PHD finger-ome identified a single putative exception to this rule, namely the first PHD finger of KDM5B (KDM5B-PHD1) that bears an Asp as the acidic wall residue, in spite of being found to bind H3 in an extended conformation based on an NMR structure of the complex (PDB: 2MNZ [47], Supplementary Figures S4A and S12). However, in that structure, the H3 peptide fits into a groove close to the N-terminus of KDM5B-PHD1 rather than running parallel to the domain as observed for the PHD fingers that bind H3 peptide in an extended conformation (Supplementary Figure S12). Such arrangement resembles the conformation observed for H3 bound to UHRF1 PHD in the co-crystal structure reported by Wang et al. [48] (Supplementary Figure S12). However, there are six other independent structures of UHRF1 PHD in complex with H3 N-terminal peptide bound in a helical fold (Supplementary Figure S12) [17,49–53]. We therefore propose that, as observed for UHRF1, the KDM5B-PHD1 can also recognize the H3 N-terminal tail in a helical fold.
We identify a subclass of 36 human PHD fingers containing an acidic wall residue Asp/Glu as potential consensus to molecular recognition of the helical H3 tail. The minimal overlap observed between the acidic wall subclass and the subclass comprising the key conserved Trp residue corresponding to specific readout of methylated-K4 suggests a level of incompatibility between these two molecular recognition features. Indeed, methylation of H3K4 was found to weaken or completely abrogate histone binding in several PHD fingers that recognize helical H3, such as BAZ2A/B [26], DPF3b [11] and MOZ [54]. This trend of incompatibility is particularly evident in the KDM5 subfamily, where, in all its members, only the first PHD finger (PHD1) has an acidic wall and this is mutually exclusive with the conserved Trp for methylated-K4 recognition that is instead present in PHD2 and PHD3 (Figure 2C). The interaction between unmodified H3 N-terminal peptide and both KDM5A-PHD1 and KDM5B-PHD1 has been recently characterized using NMR [35,36,47]. Interestingly, the patterns of CSPs observed were in both cases consistent with the one observed for BAZ2A/B PHDs (Figure 4).

The region corresponding to the acidic wall residue is often found as highly acidic, with additional Asp/Glu residues found either immediately before or after the acidic wall residue (Figure 2A–C and Supplementary Figure S5). Sequence analysis showed that 22 of these 36 PHD sequences contain at least two adjacent acidic residues (Supplementary Figure S5), suggesting the prevalence of a double acidic patch. We provide evidence that full neutralization of this double acidic patch abrogates H3 binding in BAZ2A PHD, highlighting its important role (Figure 5 and Supplementary Figure S2). We propose that the negatively charged patch at the acidic wall helps to stabilize the helical fold of H3 by forming electrostatic interactions with the positive dipole of the histone helix. The carboxylate side chain(s) at the acidic wall can help to induce the helical bound conformation in H3 by interacting with the basic side chains of K4 and Q5 at the start of the helix. In addition, it can form a salt bridge with the guanidinium group of R8, as shown by recently solved crystal structures of the DPF of DPF3b (PDB: 5I3L [55]), MOZ (PDB: 5B75 [56]) and MORF (PDB: 5U2J [57]), and thus probably occurring in BAZ2A PHD. Interestingly, mutation A275D in the DPF of MOZ just upstream of acidic wall residue D276, thus installing a double-negative charge as in BAZ2A, enhanced the binding of H3K14ac peptide by 3- to 4-fold [54].

It has been shown that PTMs can affect the secondary structure of histones [58]. It is tempting to speculate that induction or stabilization of the helicity of histone tails could represent an additional layer of regulation in epigenetic processes beyond or in cross-talk with PTMs. Within the context of tandem epigenetic reader domains, the H3 helical fold has been shown to be important for simultaneous recognition of distinct regions of the H3 tail by two epigenetic reader domains on the same protein. For example, the H3 helical conformation induced by UHRF1 PHD binding was found to be essential for productive recognition of K9 modification states by the neighboring tudor domain [49]. Similarly, the H3 helical fold was found to be critical for simultaneous recognition of K4 and K14 modifications by the double PHD finger domain of MOZ [15,54,56]. Our work suggests that, in BAZ2A/B and related proteins, the helical fold of the bound H3 N-terminal tail could facilitate productive simultaneous recognition of both unmodified K4 and downstream marks, e.g. K14ac, by the neighboring PHD finger and BRD, respectively, which warrant future investigation.

In conclusion, we propose that among the large PHD family exists a class of PHD fingers with a distinct recognition mode of the histone H3 tail that induces H3 to adopt a helical fold after K4. PHD fingers that belong to this class are characterized by the presence of a conserved Asp/Glu residue within a short acidic patch made of a helical turn or loop just before the first β-strand. We show that H3 helicity is critical for molecular recognition by this subclass of PHD fingers and identify mutations at K4 and Q5 in H3 that either enhance or weaken the binding affinity by stabilizing or disrupting the peptide helicity, respectively. This mutagenesis approach may provide a rapid and direct strategy to identify other reader domains that also recognize helical H3 tails. Our work has also implications for drug design. The growing interest in targeting epigenetic reader domains with small molecules has led to many examples of successful campaigns delivering potent chemical probes in particular for BRDs [59,60], but also for methyl-lysine reader domains such as malignant brain tumor domains [61–63] and chromodomains [63–66]. In contrast, relatively little progress has been made on targeting PHD fingers, with only two examples reporting weak-binding fragments [67] and screening-active compounds [68], suggesting low ligandability for this class of reader domains. Drug design approaches to stabilize the helical conformation, e.g. by using stapled peptides, or to mimic the helix recognition pharmacophore could provide attractive new strategies to aid the development of epigenetic chemical probes that disrupt this class of reader–histone interactions.
Experimental
Protein expression and purification
Expression and purification of BAZ2A PHD and BAZ2B PHD were performed as described recently [26]. $^{15}$N and $^{15}$N/$^{13}$C uniformly labeled BAZ2A PHD and BAZ2B PHD were expressed in modified M9 minimal medium [69] where the sole sources of nitrogen and carbon were 1 g/l $^{15}$NH$_4$Cl (Goss Scientific) and 2 g/l $^{13}$C-$d_6$-glucose (Goss Scientific) as appropriate. The expression conditions and the purification procedures used for labeled proteins were the same as for unlabeled samples [26].

Site-directed mutagenesis
Mutations were introduced into BAZ2A/B PHD fingers by polymerase chain reaction (PCR) amplification of the original construct using Phusion DNA Polymerase (Thermo Fisher Scientific) and a specific pair of primers for each mutation (Supplementary Table S4). The PCR amplification product was incubated with Dpn I (New England BioLabs) for 1 h at 37°C to digest the parental DNA strands and then used to transform *Escherichia coli* DH5α cells. Transformed cells were grown on lysogeny broth (LB) agar plates supplemented with 100 μg/ml ampicillin for 16 h at 37°C. Single colonies were picked to inoculate 5 ml of LB plus 100 μg/ml ampicillin and grown for 16 h at 37°C. The DNA was extracted from the bacterial cultures using the QIAprep Spin Miniprep Kit (Qiagen) and the presence of the desired mutation was checked by DNA sequencing. Electrospray ionization mass spectrometry analyses confirmed that the mutant constructs were successfully translated into the correctly mutated proteins.

Peptide synthesis
Peptides were synthesized by standard automated solid-phase synthesis on a ResPep SL peptide synthesizer (Intavis) using Fmoc-protected amino acids and Rink Amide resin (Novabiochem). Amino acids were coupled twice adding 1.05 equivalents of Fmoc-protected amino acid, 1 equivalent of $N,N,N'$, $N'$-tetramethyl-O-$(1H$-benzotriazol-1-yl)$uraium hexafluorophosphate and 1 equivalent of $N$-methylmorpholine with 5-fold excess over the resin. Peptides were cleaved from the resin and deprotected by incubation of the resin for 3 h with 1 ml of cleavage mixture containing 97.5% trifluoroacetic acid (TFA) and 2.5% water, leaving all peptides amidated at the C-terminus. Peptides were precipitated by the addition of 5 ml of ice-cold diethyl ether and pelleted by centrifugation. The resulting pellets were washed twice with diethyl ether. High-performance liquid chromatography (HPLC) purification was performed on a Gilson Preparative HPLC System using a Zorbax 300SB-C18 column (5 μm particle size, 250 × 9.4 mm) run at 4 ml/min. The solvents used were A (99.9% water and 0.1% formic acid or TFA) and B (94.9% acetonitrile, 5% water and 0.1% formic acid or TFA). A linear gradient from 0% to 10% B was used. All the peptides were retained during the run but were eluted before the gradient, i.e. in 100% A, except for ARTAATARKS and ARTKQTARKS, which were eluted at the beginning of the gradient. Removal of formic acid and TFA was performed using the VArriPure IPE column, and the absence of TFA was confirmed by $^{19}$F NMR. Purified peptides were submitted to liquid chromatography–mass spectrometry (LC–MS) analysis (Supplementary Figure S13). LC–MS analyses were performed with an Agilent Technologies 1200 series HPLC connected to an Agilent Technologies 6130 quadrupole spectrometer and a diode array detector. Chromatography runs were conducted with a Waters XBridge C18 column, 50 mm × 2.1 mm, with a 3.5 μm particle size with a mobile phase of water/acetonitrile +0.1% formic acid using a gradient from 95:5 to 90:7.5 min.

NMR spectroscopy
NMR spectra were acquired from $^{15}$N or $^{15}$N/$^{13}$C-labeled samples of BAZ2A PHD and BAZ2B PHD at a concentration of 350 μM in a buffer containing 50 mM KCl, 1 mM dithiothreitol (DTT), 0.02% (w/v) Na$_2$EDTA, 10% D$_2$O and 25 mM K$_2$HPO$_4$ at a pH of 6.9 for BAZ2A PHD and a pH of 6.5 for BAZ2B PHD. All NMR experiments were performed at 25°C using an AV-500 MHz Bruker spectrometer equipped with a 5 mm CTPXI 1H–$^{13}$C/$^{15}$N/D Z-GRD cryoprobe. Sequence-specific backbone assignments were obtained for BAZ2A PHD and BAZ2B PHD from the identification of intra- and inter-residue resonances in the following spectra: [[$^{15}$N,$^{1}$H]-HSQC, $^{15}$N/$^{13}$C/1H HNCO, HNCA, HN(CA)CB and HN(CO)CA]C. Acquisition times used in the [[$^{15}$N,$^{1}$H]-HSQC experiments were 120 ms for $^1$H and 60 ms for $^{15}$N. Typical acquisition times in the three-dimensional experiments were: 100 ms for $^1$H, 14–19 ms for $^{15}$N and 7–12 ms for $^{13}$C. All the NMR spectra were processed using the program TopSpin (Bruker) and analyzed using the package CcpNmr Analysis [70].
In CSP experiments, the chemical shift differences in proton (Δδ_H) and nitrogen (Δδ_N) were combined to obtain a weighted chemical shift difference (Δδ_weighted) using the following equation: Δδ_weighted = |Δδ_H| + |Δδ_N| * 0.14, where 0.14 is a scaling factor required to account for the difference in the range of amide proton and amide nitrogen chemical shifts [71]. Shifted residues were clustered based on the extent to which they showed a 0.14, where 0.14 is a scaling factor required to account for the difference in the range of amide proton and amide nitrogen chemical shifts [71]. Shifted residues were clustered based on the extent to which they showed a CSP into strong (Δδ_weighted value above the average chemical shift plus two times the standard deviation), medium (Δδ_weighted value above the average chemical shift plus the standard deviation) and weak (Δδ_weighted value above the average chemical shift). CSPs in the slow exchange regime on the NMR timescale were analyzed using the 'minimal shift approach' [72]. The chemical shift change for each backbone amide group was measured from the peak detected in the HSQC spectrum recorded on the free form to the nearest peak detected in the HSQC spectrum recorded on the bound form. Δδ_H and Δδ_N were combined as described before to obtain a minimal Δδ_weighted.

Isothermal titration calorimetry
All calorimetric experiments were performed on a MicroCal iTC200 microcalorimeter (GE Healthcare) at 25°C in a buffer containing 20 mM HEPES at a pH of 8, 150 mM NaCl and 0.5 mM Tris(2-carboxyethyl) phosphine. All ITC experiments were carried out titrating peptide solutions (1.5–3 mM) into protein solutions (80–120 μM) loaded in the calorimeter cell, performing one first injection of 0.4 μl followed by 19 injections of 2 μl. The data were analyzed using the MicroCal™ software package subtracting the data from an independent titration of peptide into buffer to account for heat of dilution, and then fitted using a single-binding site model. Protein concentration was determined by measuring absorbance at 280 nm using the following extinction coefficients: BAZ2A PHD ε_{280} = 6990 M^{-1} cm^{-1} and BAZ2B PHD ε_{280} = 8480 M^{-1} cm^{-1}. Lyophilized peptides were weighted and dissolved in an appropriate volume of buffer to obtain the desired concentration.

X-ray crystallography
Crystals of BAZ2A PHD in the apo form were grown at 18°C using the sitting drop vapor diffusion method by mixing equal volumes of protein [6 mg/ml in 20 mM Tris–HCl (pH 8), 150 mM NaCl and 2 mM DTT] and crystallization buffer (2.2 M Na/K phosphate buffer at a pH of 8.5). To obtain crystals of the BAZ2A PHD-H3 10-mer complex, preformed apo BAZ2A PHD crystals were transferred and soaked overnight into a solution containing 2 mM H3 10-mer (ARTKQTARKS) in crystallization buffer and cryoprotected in 1.6 mM H3 10-mer, 1.7 M Na/K phosphate and 20% glycerol. The data sets were collected at the beamline ID29 at European Synchrotron Radiation Facility and processed with XDS [73,74] and AIMLESS [75], to 2.4 Å of resolution. The structure of the complex was determined by isomorphous replacement with the apo form of BAZ2A PHD (PDB entry: 4QF2 [26]). Manual model building and refinement were carried out using Coot [76] and Refmac5 [77]. The quality of the models was checked by MolProbity [78], and all structure figures were generated using PyMOL (The PyMOL Molecular Graphics System, Version 1.7.05, Schrödinger, LLC).

Molecular dynamics
System set-up
The X-ray crystal structure of BAZ2A PHD in complex with H3 10-mer (ARTKQTARKS) was used as the starting structure of the corresponding simulation. The missing residues K9 and S10 were added choosing a suitable low-energy rotamer from PyMOL and minimized for 4000 steps with the rest of the protein fixed. The initial structures of the H3 10-mer with K4A/Q5A and K4G/Q4G mutations to generate ARTATARKS and ARTGGTARKS, respectively, were built from the WT structure, with the point mutations performed in PyMOL. The structures of the H3 peptides obtained this way were used to simulate the peptides in the unbound state, i.e. in aqueous solution, as well. All models were solvated in a TIP3P water box with a padding of 15 Å from the edge of the box to any protein atom. The system charges were neutralized with sodium or chloride ions as appropriate.

Simulation protocol
MD simulations were carried out using the NAMD program [79] and the CHARMM 36 force field [80]. Initially, the solvated systems were minimized for 3000 steps with the protein restrained to eliminate residue unfavorable interactions between the protein and the solvent, followed by another 5000 steps with all atoms free to move. Heating of the systems from 0 to 300 K was achieved in 100 ps (time step of 1 fs), with fixed
protein backbone atoms to allow relaxation of the solvent. The systems were subsequently equilibrated for 600 ps (time step of 2 fs) with all atoms free to move. The NPT ensemble was used during the production simulations, which involved four replicates of 80 ns each (time step of 2 fs). The temperature was controlled with a Langevin thermostat at 300 K, and the pressure with a Nose–Hoover Langevin piston barostat at 1 bar. A SHAKE constraint was applied to all bonds containing hydrogen atoms. Short-range non-bonded interactions were switched at 10 Å and cut off at 12 Å, and particle mesh Ewald summation was employed for long-range non-bonded interactions. Consistency and stability throughout the MD replicas were assessed (Supplementary Table S5). The per-residue secondary structure calculation was performed using the Timeline plugin v.2.3 and the hydrogen-bond contacts with the HBonds plugin v.1.2, both contained in VMD v. 1.9.2 [81]. Pair-wise distribution differences among simulated systems were assessed statistically using the two-tailed Mann–Whitney U-test, as implemented in the statistical package R [82].

Sequence alignment

Sequences of domains that belong to the PHD family and whose structures were solved in complex with an H3 N-terminal tail peptide were identified with the software Dali [83] using as input the structure of BAZ2A PHD (4QF2). The sequences of human PHD fingers were obtained from the Structural Genomic Consortium database [1]. The multiple sequence alignment was performed using MAFFT (Multiple Alignment using Fast Fourier Transform) [84] and analyzed using Jalview [85].

Circular dichroism

CD spectra were acquired from H3-derived peptides dissolved in water (30 μM) at increasing concentrations of TFE using a Bio-Logic CD spectrometer with a cuvette with a path length of 1 mm, at a temperature of 20°C, with a bandwidth of 0.5 nm and a sampling time of 0.5 s. Each spectrum represents the average of three accumulations minus the signal from the blank. Additionally, a constant was added or subtracted to CD spectra so that ellipticity at high wavelengths was 0. Spectra deconvolution was performed using the CONTIN algorithm [86] implemented into DichroWeb [87].

Accession numbers

The atomic coordinates and structure factors have been deposited in the PDB with the accession number: PDB ID: 5T8R (BAZ2A PHD in complex with H3 10-mer). NMR assignments for BAZ2A PHD and BAZ2B PHD have been deposited in the BMRB with deposition numbers 26754 and 25988, respectively.

Abbreviations

BAZ, bromodomain adjacent to zinc finger; BMRB, biological magnetic resonance bank; BRD, bromodomain; CD, circular dichroism; CSP, chemical shift perturbation; DPF, double PHD finger; DTT, dithiothreitol; HPLC, high-performance liquid chromatography; HSQC, heteronuclear single-quantum coherence; ITC, isothermal titration calorimetry; LB, lysogeny broth; LC–MS, liquid chromatography–mass spectrometry; MD, molecular dynamics; NoRC, nucleolar remodeling complex; PCR, polymerase chain reaction; PDB, Protein Data Bank; PHD, plant homeodomain; PTM, post-translational modification; rDNA, ribosomal DNA; RMSD, root mean square deviation; SNF2h, sucrose nonfermenting protein 2 homolog; TFA, trifluoroacetic acid; TFE, 2,2,2-trifluoroethanol; WT, wild type.

Author Contribution

A.C. supervised the project. A.B., A.A., X.L. and A.C. designed the research. A.B., A.A., X.L. and M.B. performed the research. A.B., A.A., X.L. and A.C. analyzed data. A.B., X.L. and A.C. wrote the paper with contribution from the other authors.

Funding

This work was supported by the U.K. Biotechnology and Biological Sciences Research Council (BBSRC, grants BB/J001201/2 and BB/G023123/2, David Phillips Fellowship to A.C.), the European Research Council ERC-2012-StG-311460 DrugE3CRLs (Starting Grant to A.C.) and the Wellcome Trust (Strategic Awards 100476/Z/12/Z for biophysics and drug discovery, and 094090/Z/10/Z for structural biology and X-ray crystallography to the Division of Biological Chemistry and Drug Discovery). A.B. was partly funded by a Foundation Blancefor Boncompagni-Ludovisi Scholarship. A.A. is funded by a BBSRC PhD Studentship. X.L. is a Marie
Skłodowska-Curie Actions Individual Fellow [H2020-MSCA-IF-2015-806323]. M.B. was funded by a James Black Prize Vocational Scholarship.

Acknowledgements
We thank the European Synchrotron Radiation Facility (ESRF) for beamtime (proposal mx-1692) and support at beamline ID29, the Division of Computational Biology of the University of Dundee for providing computing cluster time, Morgan Gadd for assistance with X-ray data collection and refinement, and Conny W.H. Yu and Uli Schwarz-Linek at the University of St Andrews for access to circular dichroism spectrometers.

Competing Interests
The Authors declare that there are no competing interests associated with the manuscript.

References