Crystal structure of the proximal BAH domain of the polybromo protein

Antony W. OLIVER^{*1}, Sarah A. JONES^{*}, Stephen Mark ROE^{*}, Steve MATTHEWS[†], Graham H. GOODWIN[‡] and Laurence H. PEARL^{*}

^{*}CR-UK DNA Repair Enzymes Group, Section of Structural Biology, The Institute of Cancer Research, 237 Fulham Road, Chelsea, London SW3 6JB, U.K., [†]Department of Biological Sciences and Centre for Structural Biology, Imperial College of Science, Technology and Medicine, Wolfson Laboratory, Exhibition Road, London SW7 2AY, U.K., and [‡]Department of Molecular Carcinogenesis, The Institute of Cancer Research, Haddow Laboratories, 15 Cotswold Road, Sutton SN2 5NG, Surrey, U.K.

The BAH domain (bromo-associated homology domain) was first identified from a repeated motif found in the nuclear protein polybromo – a large (187 kDa) modular protein comprising six bromodomains, two BAH domains and an HMG box. To date, the BAH domain has no ascribed function, although it is found in a wide range of proteins that contain additional domains involved in either transcriptional regulation (e.g. SET, PHD and bromo-domain) and/or DNA binding (HMG box and AT hook). The molecular function of polybromo itself also remains unclear, but it has been identified as a key component of an SWI/SNF (switching/sucrose non-fermenting)-related, ATP-dependent chromatin-re-

modelling complex PBAF (polybromo, BRG1-associated factors; also known as SWI/SNF-B or SWI/SNF β). We present in this paper the crystal structure of the proximal BAH domain from chicken polybromo (BAH1), at a resolution of 1.6 Å (1 Å = 0.1 nm). Structure-based sequence analysis reveals several features that may be involved in mediating protein–protein interactions.

Key words: BAH domain, chromatin remodelling, polybromo, ORC1, RSC, SWI-SNF.

INTRODUCTION

In a search for transcription factors that contained bromodomains (domains that interact with acetylated lysines of histones H3 and H4), we cloned the cDNA of a novel chicken protein, polybromo [1]. This ubiquitously expressed 187 kDa protein has six bromodomains plus two copies of a novel domain that we named the BAH domain (bromo-associated homology domain).

The precise biological role of polybromo is not fully understood, but the human orthologue (BAF180) has been identified as a component of the SWI/SNF (switching/sucrose non-fermenting) chromatin-remodelling complex PBAF (polybromo, BRG1-associated factors), a complex that has been shown to stimulate transcription of chromatin templates *in vitro* [2].

It is known that several sequence-specific DNA-binding proteins, including nuclear hormone receptors, can recruit the multiprotein Mediator complex [also termed TRAP/SMCC, where TRAP stands for thyroid-hormone-receptor-associated protein and SMCC for SRB (suppressor of RNA polymerase B)- and MED (Mediator)-containing cofactor complex] along with RNA polymerase II to a promoter to form an initiation complex. Reconstituted *in vitro* chromatin transcription assays have demonstrated that purified human PBAF stimulates Mediator/TRAP/SMCC-activated transcription with some nuclear hormone receptors and sequence-specific DNA-binding proteins [3]. The closely related SWI/SNF chromatin-remodelling complex BAF (SWI/SNF-A), which does not contain polybromo, did not activate transcription in this assay.

The PBAF complex has been demonstrated to be related to the yeast RSC chromatin-remodelling complex [2]. Interestingly, there is no direct homologue of polybromo in yeast that is contained within a single polypeptide chain, instead its component

domains appear to be distributed among three separate proteins: RSC1, RSC2 and RSC4. Both RSC1 and RSC2 consist of a single BAH domain, two bromodomains, plus a DNA-binding AT-hook, whereas RSC4 contains just two bromodomains [4]. The RSC complex is involved in the transcriptional activation and repression of a wide variety of yeast genes [5–7].

It is likely therefore that the polybromo and RSC1/RSC2/RSC4 proteins are involved in transcriptional regulation through their association with their respective PBAF–RSC complexes. The RSC complex is also involved in chromosome segregation and kinetochore function, with the BAH domain of RSC2 shown to be absolutely essential for 2 μ m plasmid maintenance in yeast, a process dependent on chromatin remodelling events [8]. The BAH domains in polybromo are likely to be involved in mediating interactions with other components of the complex, and/or with DNA-binding proteins, to recruit polybromo-containing complexes to target promoters/chromosome cohesion sites.

Apart from polybromo and its homologues, BAH domains have been identified in a range of other DNA and chromatin-associated proteins, including the transcription factor Ash1, the CpG-DNA methylase DNMT1 [DNA (cytosine-5)-methyltransferase 1] and the replication origin recognition protein ORC1 (origin recognition complex subunit 1). Sequence alignments of the BAH domains from various proteins have identified a conserved region of approx. 130 amino acids, but with large insertions of variable length in two regions of the domain, in some examples [9,10]. The BAH domains of polybromo, the RSC proteins and Ash1 are closely related but appear to be only loosely related to the BAH domains of the ORC1 and DNMT1 proteins. To gain further insight into the structure and function of polybromo, we have now determined the crystal structure of the proximal (more N-terminal) BAH domain from chicken polybromo.

Abbreviations used: BAH domain, bromo-associated homology domain; DNMT1, DNA (cytosine-5)-methyltransferase 1; DTT, dithiothreitol; HSQC, heteronuclear single-quantum coherence; ORC1, origin recognition complex subunit 1; PBAF, polybromo, BRG1-associated factors; SWI/SNF, switching/sucrose non-fermenting.

¹ To whom correspondence should be addressed (email antony.oliver@icr.ac.uk).

Co-ordinates of the protein sequence used will appear in the Protein Data Bank under the accession number 1W4S.

Table 1 Data collection and refinement parameters

The values given in parentheses represent highest shell values. Se-Met, seleno-methionine; $I/\sigma(I)$, the average intensity in the resolution bin divided by the S.D. of that average; $R_{\text{merge}} = (\sum_{hkl} |I_{hkl} - \langle I_{hkl} \rangle|) / (\sum_{hkl} I_{hkl})$, where I_{hkl} is the intensity for each observation/symmetry equivalent of a particular reflection, hkl , and $\langle I_{hkl} \rangle$ is the average intensity over all observations/symmetry equivalents of the reflection; $R_{\text{cryst}} = (\sum_{hkl} |F_{\text{obs}} - F_{\text{calc}}|) / (\sum_{hkl} |F_{\text{obs}}|)$, where F_{obs} are the observed structure factor amplitudes and F_{calc} those calculated from the model.

	Native	Se-Met (peak)	Se-Met (inflection)	Se-Met (remote)
Data collection				
Wavelength (Å)	0.9330	0.9787	0.9793	0.9724
Spacegroup	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁		
Unit cell (Å)	$a = 39.60, b = 57.71, c = 70.38$	$a = 39.73, b = 57.49, c = 69.96$		
Resolution range (Å)	29.62–1.55 (1.63–1.55)	44.41–2.7 (2.85–2.7)	57.73–2.7 (2.85–2.7)	57.74–2.7 (2.85–2.7)
R_{merge}	0.073 (0.258)	0.080 (0.151)	0.066 (0.128)	0.086 (0.187)
$I/\sigma(I)$	17.9 (4.7)	26.5 (15.6)	30.7 (17.6)	25.9 (13.8)
Completeness (%)	99.4 (100)	100 (100)	100 (100)	100 (100)
Multiplicity	4.7 (4.7)	13.3 (14.2)	13.4 (14.2)	13.1 (14.2)
No. of observations	111 577 (16 472)	63 391 (9570)	63 766 (9360)	62 453 (9313)
No. of unique reflections	23 958 (3467)	4754 (674)	4759 (669)	4760 (670)
Structure refinement				
No. of atoms (protein)	1221			
No. of atoms (all)	1327			
R_{cryst}	0.200			
R_{free}	0.230			

MATERIALS AND METHODS

NMR spectroscopy

¹⁵N-labelled samples of BAH1A or BAH1 were produced from bacteria grown in ECO-E.Coli-OD2 N media (Silantes, München, Germany). NMR samples of the protein were prepared to a final concentration of approx. 250 µM and a volume of 500 µl in 50 mM acetate buffer (pH 5.0), 100 mM sodium chloride and 10 % ²H₂O. One-dimensional ¹H and two-dimensional ¹H-¹⁵N HSQC (heteronuclear single-quantum coherence) NMR spectra of ¹⁵N-labelled BAH1A and BAH1 domains were performed on a 500 MHz four-channel Bruker DRX500 spectrometer equipped with a z-shielded gradient triple resonance cryoprobe.

Expression and purification

Escherichia coli BL21(DE3) pLysS cells (Novagen, Madison, WI, U.S.A.) were transformed with the plasmid construct pET16b-BAH1 that encodes amino acids 932–1103 of the *Gallus gallus* polybromo protein, preceded by an N-terminal Factor Xa cleavable His₁₀ tag.

A 250 ml flask containing 50 ml of L-broth (1 %, w/v, tryptone, 0.5 %, w/v, NaCl and 0.5 %, w/v, yeast extract), supplemented with 50 µg/ml carbenicillin and 34 µg/ml chloramphenicol, was inoculated with a single transformed bacterial colony. This was grown, in an orbital incubator set at 37 °C and 225 rev./min, until the attenuation (D_{600}) of the culture reached 0.6. This 'starter' culture was then stored overnight at 4 °C.

A starter culture of 10 ml was used to inoculate a 2 litre flask containing 1 litre of L-broth supplemented as before with antibiotics. The culture was grown in an orbital incubator set at 37 °C and 200 rev./min. Once the D_{600} of the cell culture had reached 0.6, expression of the BAH1 protein was induced by the addition of isopropyl β-D-thiogalactoside to a final concentration of 1 mM. Cultures were grown for a further 3 h, when the cells were harvested by centrifugation (5000 g, 10 min, 10 °C). Cell pellets were stored at –70 °C until required.

The cell pellet resulting from 2 litre of culture was resuspended, on ice, in 40 ml of buffer A [50 mM Hepes, pH 7.5, 500 mM NaCl, 5 mM imidazole and 5 %, v/v, glycerol, supplemented with

protease inhibitors (Roche, Lewes, East Sussex, U.K.)]. Cells were then lysed by sonication (15 × 5 s bursts, on ice, at 75 % amplitude, large parallel probe; Jencons Ultrasonic Processor), after which cell debris and insoluble materials were removed by centrifugation (40 000 g, 60 min at 10 °C).

The supernatant from this step was filtered using 1 µm Whatman filter and then applied to an IMAC column (Talon; BD Biosciences Clontech UK, Oxford, U.K.) equilibrated in buffer A. Bound protein was eluted with a linear gradient from buffer A to buffer B (50 mM Hepes, pH 7.5, 500 mM NaCl, 300 mM imidazole and 5 % glycerol).

Fractions containing BAH1 were identified by SDS/PAGE, pooled, concentrated and then dialysed several times against buffer C [50 mM Hepes, pH 7.5, 500 mM NaCl, 1 mM EDTA, 1 mM DTT (dithiothreitol) and 5 % (v/v) glycerol]. To facilitate cleavage of the N-terminal His tag by Factor Xa, CaCl₂ was added to the dialysate to a final concentration of 1 mM. Approximately 150 units of Factor Xa were added, and the cleavage reaction was allowed to proceed at room temperature (22 °C) for an average of 3 h; progress of the reaction was monitored by SDS/PAGE analysis.

Once the tag had been fully cleaved, Factor Xa was removed by application of the sample to a benzamide affinity column (Amersham Biosciences, Little Chalfont, Bucks., U.K.). The flow-through resulting from this column was retained, concentrated and then applied to a Sephadex 75 size-exclusion column (Amersham Biosciences) equilibrated in buffer D [10 mM Hepes, pH 7.5, 500 mM NaCl, 1 mM DTT, 1 mM EDTA and 5 % (v/v) glycerol].

Fractions corresponding to BAH1 were again identified by SDS/PAGE, pooled and then concentrated to 11.5 mg/ml utilizing Vivaspin 6 and 500 concentrators (10 kDa cut-off). The protein was, at this stage, judged to be at > 98 % purity, and thus suitable for crystallization trials.

Production of seleno-methionine labelled protein

The plasmid construct pET16b-BAH1 was transformed into the *E. coli* strain B834 (DE3), a methionine auxotroph suitable for facilitating seleno-methionine incorporation. The methodology outlined by the EMBL-Heidelberg Protein Expression and

Purification Facility was used, detailed in their Protocol Database (<http://www.embl-heidelberg.de/ExternalInfo/geerlof/draft-frames>). Seleno-methionine labelled protein was purified and crystallized following the same procedures as those for the native protein.

Crystallization and data collection

Crystallization trials were performed at 11.5 mg/ml in hanging-drop experiments using Structure Screen I [MDL (Molecular Dimension Limited), Soham, Cambs., U.K.]. Large crystals were observed, almost immediately, under two conditions: (i) 0.1 M Hepes (pH 7.5) and 1.5 M lithium sulphate (MDL 27) and (ii) 0.1 M Tris/HCl (pH 8.5) and 2.0 M ammonium sulphate (MDL 32).

This condition was optimized, again in hanging-drop experiments, by mixing 1 μ l of protein (7.5 mg/ml in 10 mM Hepes, pH 7.5, 500 mM NaCl, 1 mM DTT and 1 mM EDTA) with 1 μ l of precipitant containing 0.1 M Tris/HCl (pH 8.0), 20% (w/v) sucrose and a range of 0.6–1.6 M ammonium sulphate. The crystals were cryoprotected by swiping through successive buffers that contained increasing amounts of sucrose, to a final concentration of 40% (w/v).

Data up to 1.6 Å (1 Å = 0.1 nm) were collected from a single crystal at 100 K at the ESRF (European Synchrotron Radiation Facility, Grenoble, France) and recorded on a Mar charge-coupled-device camera. Images were integrated using MOSFLM [15] and reduced/scaled using programs of the CCP4 suite [16].

The protein crystallized in the spacegroup $P2_12_12_1$ had cell dimensions of $a = 39.60$ Å, $b = 57.71$ Å and $c = 70.37$ Å. Statistics for the data collection are given in Table 1

Structure determination and refinement

Attempts to solve the structure using molecular replacement methods failed to provide a solution. Seleno-methionine labelled protein was therefore produced to provide independent phasing information.

Seleno-methionine labelled protein crystallized under the same conditions as the native protein, and the crystals produced were isomorphous. Data were collected up to 2.7 Å at the ESRF. Initial phases were calculated from a MAD experiment using SHARP [17,18], and a preliminary model containing 138 residues automatically built by ARP/wARP [19].

Difference maps were used to extend and rebuild the initial model using O [20] and Coot [21]. Iterative cycles of refinement (REFMAC5 [15]) and manual intervention produced the current model that consists of 1221 protein atoms, one chloride ion and 105 solvent atoms, with $R = 0.200$ and $R_{\text{free}} = 0.230$ for 5% of the data omitted from refinement. Co-ordinates have been deposited in the Protein Data Bank (PDB no. 1W4S).

RESULTS

It was not possible to predict accurately the exact N- and C-terminal boundaries of the polybromo proximal BAH domain, even using the existing structure of the BAH-containing domain of ORC1 as a guide (PDB no. 1M4Z), due to the large amino acid sequence divergence between the two proteins.

We therefore cloned an extended fragment that included the previously identified BAH homology region [9,10] plus additional flanking sequences on either side (Figure 1). The resultant protein (construct BAH1A: amino acids 878–1115) was expressed in bacteria and then extracted with a non-denaturing buffer (15 mM sodium phosphate, pH 7.6, and 0.5 M NaCl). It was then purified

Figure 1 Schematic representation of the functional domains of chicken polybromo

The BAH1A construct is represented by the top line (amino acids 878–1115). The BAH1 construct is represented by the lower line (amino acids 932–1103). BROMO, bromodomain; BAH, BAH domain; HMG, truncated HMG box.

Figure 2 NMR spectroscopic analysis of protein constructs

(a) Methyl group region from the one-dimensional ^1H NMR spectra of BAH1A (top) and BAH1 (bottom). Unstructured methyl group resonances are indicated by an arrow. (b) Two-dimensional ^1H - ^{15}N HSQC of ^{15}N -labelled BAH1. The highly dispersed resonances in the two-dimensional ^1H - ^{15}N HSQC spectrum clearly indicate a fully folded species.

from the cell extract on a Ni-affinity column, facilitated by a vector-encoded His tag (see the Materials and methods section).

One-dimensional ^1H and two-dimensional ^1H - ^{15}N NMR spectra of ^{15}N -labelled BAH1A showed that the protein contained a

Figure 3 Structure of the proximal BAH domain of chicken polybromo

(a) Stereo-pair secondary-structure cartoon, coloured blue to red from the visible N-terminus at residue 956 to the visible C-terminus at residue 1099. (b) As in (a), but rotated 90° around the horizontal axis. All molecular graphic images in this and subsequent Figures were generated using PyMOL [W. L. DeLano (2002) The PyMOL Molecular Graphics System, see <http://www.pymol.org>].

globular domain, but line widths indicated that approx. 30–40 amino acids were highly flexible and unstructured (Figure 2a). Trypsinization of the 28 kDa BAH1A protein resulted in a metastable product of approx. 18 kDa, corresponding to the size predicted for the globular component from the NMR data. On the basis of these analyses, a second shorter construct was expressed (BAH1: amino acids 932–1103), in which the highly charged regions at the two termini were omitted. Furthermore, one-dimensional ^1H and two-dimensional ^1H - ^{15}N NMR analyses on this construct confirmed that the unstructured regions had been removed and the protein was substantially globular, representing a much more tractable target for structural investigation (Figure 2b).

Crystal structure of the proximal BAH domain of polybromo

The purified BAH1 protein was crystallized and the crystal structure was determined at 1.6 Å resolution using the multiwavelength anomalous diffraction method, with phases provided from selenomethionine labelled protein. At the N-terminus of the protein, amino acids 932–954 have no readily discernable electron density and appear to be disordered. At the C-terminus, amino acids

1101–1103 are similarly disordered. The polybromo-BAH1 structure comprises a single globular domain consisting predominantly of β -strands (Figures 3 and 4). The core of the protein fold is created from an open, distorted barrel comprising β -strands 1, 2, 3, 5, 6 and 8, which is interrupted by a 3_{10} helix (B) positioned between β -strands 5 and 6 (Figure 4). The barrel is extended on one side by continuation of the β -sheet by strand 9. Despite the apparently open structure of this barrel, there is no solvent channel running through it, as it is completely filled with hydrophobic residues.

Further elaboration of the fold is provided by a 3_{10} helix (A) and a β -strand (4) that together create an extended loop which projects out from the β -barrel to connect strands 3 and 5. A mixture of short helices and β -strands completes the rest of the fold. Of particular interest is a long C-terminal extension comprising β -strands 10 and 11, which effectively ‘wraps’ around the outside of the protein to stabilize two extended loops through the formation of antiparallel β -sheets (i.e. those between strands 4 and 11, plus 7 and 10). This segment appears to add considerable stability to the overall fold, as constructs made without this C-terminal region are insoluble (results not shown).

Figure 4 Structure-guided sequence alignment of the proximal BAH domain (BAH1) and the BAH-domain containing region of yeast Orc1p (ORC1)

(a) For BAH1, β -strands are numbered consecutively and α /3₁₀ helices are labelled alphabetically from the N-terminus to the C-terminus. Secondary-structure labelling for ORC1 follows that of Zhang et al. [11] apart from 3₁₀ helices X and Y; these are additional elements assigned after a DSSP assessment of the secondary structure. Residues invariant between the two structures are underlined and shaded dark grey, whereas conserved residues are in italics and shaded light grey. (b) Secondary-structure cartoon representations of BAH1 and ORC1. (c) Topology diagrams detailing the folds of both structures. Triangles represent β -strands and circles α -helices/3₁₀ turns. Labelling of secondary-structure elements is as for (a, b). Secondary-structure elements that make up the 'core' BAH domain are shaded black in (a) and dark grey in (b, c).

Structural comparison between the polybromo and ORC1 BAH domains

The structure of the BAH-containing domain from yeast ORC1 has previously been determined [11] (PDB no. 1M4Z). Compari-

son of the polybromo and ORC1 BAH domain structures (Figure 4) clearly shows a common 'core' fold, essentially the aforementioned distorted β -barrel. In addition, the 3₁₀ helices A and B (in BAH1) and X and Y (in ORC1) are apparently conserved between the two structures, as are β -strands 8 and 9 (BAH1)/9 and 10

Figure 5 Multiple amino acid sequence alignment of BAH domains

The alignment is subdivided into three groups (indicated by the numerals at the left-hand side of the alignment). Group 1 contains only proximal polybromo-BAH domains, group 2 the polybromo-related proteins RSC1, RSC2 and the transcription factor Ash1, with group 3 containing distal polybromo-BAH domains. Residues conserved within a group are shaded light grey; residues conserved across all groups are boxed and highlighted by a triangle; residues invariant across all groups are shown in white with a black background and further highlighted with a star. Sequence numbering and secondary-structure elements for polybromo-BAH1 are shown. Alignments were generated using Multalin [22].

(ORC1). Polybromo-BAH1 and ORC1 differ quite dramatically, however, outside these regions. The BAH domain in ORC1 is interrupted by a small domain insertion, termed the 'H' domain, which sits between the 3_{10} helix X and β -strand 7 (equivalent to 3_{10} helix A and β -strand 5 in BAH1). This insertion does not occur in polybromo-BAH1, instead a β -strand (4) serves to connect the two elements. A further α -helical section also occurs at the C-terminus of ORC1 (F and G) and again this is absent from the polybromo-BAH1 structure.

Conserved amino acids identified by sequence alignment

Amino acid sequences taken from vertebrate polybromo-BAH domains were used to seed a preliminary alignment. This was subsequently 'evolved' by stepwise addition of additional BAH domain sequences identified across a range of species and proteins (Figure 5). Alignments that included sequences from the proteins DNMT1, ORC1 and Sir3 failed to produce significant scores,

suggesting that these domains are diverged from the canonical BAH domains found in polybromo.

The alignment clearly identifies three residues: Pro⁹⁹⁹, Phe¹⁰¹⁰ and Glu¹⁰¹⁴, which are absolutely conserved in all canonical BAH domains. On examination of the polybromo-BAH1 structure, these residues were found to be clustered at a single site (Figure 6a). Pro⁹⁹⁹ and Phe¹⁰¹⁰ pack together in a face-to-edge arrangement, with the face of Phe¹⁰¹⁰ against the edge of Pro⁹⁹⁹. The side chain of Glu¹⁰¹⁴ is exposed to the solvent, sitting at the bottom of a small depression or pocket (Figure 6b).

Sequence conservation scores generated from the multiple alignments were used to produce the coloured surface representations shown in Figure 6. By comparing residues conserved within proximal BAH domains (Figures 6c and 6d) with those conserved across all BAH domains (Figures 6e and 6f), two features are highlighted, which involve the conserved amino acid triad of Pro⁹⁹⁹, Phe¹⁰¹⁰ and Glu¹⁰¹⁴. The first of these is a shallow groove or channel running along one face of the domain (Figures 6c and

Figure 6 Conserved amino acids within canonical BAH domains

(a) Secondary-structure cartoon highlighting the conserved amino acid triad Pro⁹⁹⁹, Phe¹⁰¹⁰ and Glu¹⁰¹⁴. (b) As in (a), but with a superimposed calculated molecular surface representation. The small pocket containing Glu¹⁰¹⁴ can be seen. (c–f) Coloured surface representations of polybromo-BAH1. The surfaces are coloured according to amino acid sequence conservation scores: from red (absolutely conserved), through yellow and green, to blue (not conserved). Panels (c, d) use conservation scores calculated from alignments of polybromo-BAH1 domains (group 1 in Figure 5), whereas panels (e, f) use scores calculated from alignments encompassing a wider range of proteins that contain BAH domains (groups 1–3 in Figure 5). Panels (c, e) are on the opposite face of the protein to panels (d, f). Key conserved residues are highlighted throughout.

6e), with Pro⁹⁹⁹ and Phe¹⁰¹⁰ forming part of the base of this feature. On the opposite face of the protein, the carboxylate head-group of Glu¹⁰¹⁴ forms the base of a small pocket, lined by conserved residues.

The distal BAH domain of polybromo

Polybromo contains a tandem pair of BAH domains (Figure 1). Using structure-based sequence alignment, it is evident that more C-terminal BAH domain (BAH2) should adopt the same overall fold as that described for polybromo-BAH1 (Figure 5). All major secondary-structure elements would appear to be conserved, with the notable exception of the C-terminal β -strands 10 and 11; these most probably form part of a linker region that serves to connect the two BAH domains.

DISCUSSION

There have been a number of attempts to define (and redefine) the BAH domain in terms of its secondary-structure composition. We originally identified it as an 86 amino acid repeated motif in the chicken protein polybromo [1]. Callebaut et al. [10] subsequently expanded this definition to include two additional conserved sequence regions, which correspond to β -strands 1 and 2 in our structure (Figure 3). They also showed that sequence insertions of variable lengths could occur between β -strands 1 and 2 and also between 3_{10} helix A and β -strand 5.

On the basis of the crystal structure of the BAH-containing region of yeast ORC1 protein [11], the BAH definition was further extended to include two additional β -strands (β 1 and β 2; see Figure 4b). Both of these secondary-structure elements are absent

from polybromo-BAH1, as is β -strand 3; no observable electron density is seen before residue 955, despite the expression construct having an additional 22 amino acids at the N-terminus (932–954). It is possible that these residues could adopt an antiparallel β -sheet conformation, as in ORC1, but secondary-structure predictions for this region strongly suggest that it is instead α -helical or non-structured (results not shown). We therefore suggest that the BAH domain is best represented by the following elements (shaded black in Figure 4a): a six-stranded open β -barrel core, interrupted by a 3_{10} helix (strands 1, 2, 3, 5, 6 and 8, helix B) plus a further 3_{10} helix and β -strand (A and 9). Our definition essentially agrees with that of Callebaut et al. [10].

In the last few years, a plethora of domains have been identified in proteins involved in chromatin manipulation, which are found to interact with specific post-translational modifications of histone (and other) protein side chains and mediate the so-called 'histone code' [12]. For example, bromodomains mediate interactions with acetylated lysine [13], whereas some chromodomains specifically bind sequences with methylated lysines [14]. The close association of BAH domains with bromodomains in proteins such as polybromo, RSC1 and RSC2 suggests that they might also function in 'state-specific' interactions with histones and other chromatin proteins, but we have yet to identify such an interaction experimentally.

Although an experimentally determined interaction mediated by a canonical BAH domain remains to be defined, analysis of residue conservation among BAH domains across a broad evolutionary range clearly indicates areas of the protein that are probably involved in such a function. Most striking is the absolutely conserved amino acid triad of residues Pro⁹⁹⁹, Phe¹⁰¹⁰ and Glu¹⁰¹⁴. These residues are part of distinct grooves/depressions whose dimensions would allow them to interact with segments of polypeptide from a putative partner protein. The suggestion that Glu¹⁰¹⁴ may be functionally relevant is strongly supported by mutational analysis of the polybromo-related yeast protein RSC2, where single site mutations were found to disrupt correct daughter cell segregation of the 2 μ m plasmid [8]. Of the several mutations identified, one was a single base substitution from glycine to alanine residue at position 1402. This results in a Glu $\rightarrow$ Lys mutation in the BAH domain of the protein at the equivalent residue to Glu¹⁰¹⁴ in the polybromo-BAH1 domain structure (Figure 5). While the longer lysine side chain could be accommodated without disrupting the overall folding of the RSC2 BAH domain, the charge reversal and severe alteration of hydrogen-bonding potential would certainly disrupt protein–protein interactions mediated by this face of the domain, consistent with the observed loss of function. Experiments to identify interacting proteins are in progress and will be the key to understanding BAH domain function in chromatin remodelling.

We are grateful to the ESRF for access to their synchrotron radiation facility. This work was supported by Cancer Research UK (L. H. P. and G. H. G.) and BBSRC (S. M.).

REFERENCES

- Nicholas, R. and Goodwin, G. (1996) Molecular cloning of polybromo, a nuclear protein containing multiple domains including five bromodomains, a truncated HMG-box, and two repeats of a novel domain. *Gene* **175**, 233–240.
- Xue, Y., Canman, J. C., Lee, C. S., Nie, Z., Yang, D., Moreno, G. T., Young, M. K., Salmon, E. D. and Wang, W. (2000) The human SWI/SNF-B chromatin-remodeling complex is related to yeast rsc and localizes at kinetochores of mitotic chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13015–13020.
- Lemon, B., Inouye, C., King, D. S. and Tjian, R. (2001) Selectivity of chromatin-remodelling cofactors for ligand-activated transcription. *Nature (London)* **414**, 924–928.
- Cairns, B. R., Schlichter, A., Erdjument-Bromage, H., Tempst, P., Kornberg, R. D. and Winston, F. (1999) Two functionally distinct forms of the RSC nucleosome-remodeling complex, containing essential AT hook, BAH, and bromodomains. *Mol. Cell* **4**, 715–723.
- Swanson, M. J., Qiu, H., Sumibcay, L., Krueger, A., Kim, S. J., Natarajan, K., Yoon, S. and Hinnebusch, A. G. (2003) A multiplicity of co activators is required by Gcn4p at individual promoters *in vivo*. *Mol. Cell. Biol.* **23**, 2800–2820.
- Ng, H. H., Robert, F., Young, R. A. and Struhl, K. (2002) Genome-wide location and regulated recruitment of the RSC nucleosome-remodeling complex. *Genes Dev.* **16**, 806–819.
- Yukawa, M., Koyama, H., Miyahara, K. and Tsuchiya, E. (2002) Functional differences between RSC1 and RSC2, components of a for growth essential chromatin-remodeling complex of *Saccharomyces cerevisiae*, during the sporulation process. *FEMS Yeast Res.* **2**, 87–91.
- Wong, M. C., Scott-Drew, S. R., Hayes, M. J., Howard, P. J. and Murray, J. A. (2002) RSC2, encoding a component of the RSC nucleosome remodeling complex, is essential for 2 micron plasmid maintenance in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **22**, 4218–4229.
- Goodwin, G. and Nicholas, R. (2001) The BAH domain, polybromo and the RSC chromatin remodelling complex. *Gene* **268**, 1–7.
- Callebaut, I., Courvalin, J. C. and Mornon, J. P. (1999) The BAH (bromo-adjacent homology) domain: a link between DNA methylation, replication and transcriptional regulation. *FEBS Lett.* **446**, 189–193.
- Zhang, Z., Hayashi, M. K., Merkel, O., Stillman, B. and Xu, R. M. (2002) Structure and function of the BAH-containing domain of Orc1p in epigenetic silencing. *EMBO J.* **21**, 4600–4611.
- Jenuwein, T. and Allis, C. D. (2001) Translating the histone code. *Science* **293**, 1074–1080.
- Zeng, L. and Zhou, M. M. (2002) Bromodomain: an acetyl-lysine binding domain. *FEBS Lett.* **513**, 124–128.
- Nielsen, P. R., Nietlispach, D., Mott, H. R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A. G., Murzina, N. V. and Laue, E. D. (2002) Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature (London)* **416**, 103–107.
- Leslie, A. (1995) MOSFLM Users Guide, MRC Laboratory of Molecular Biology, Cambridge, U.K.
- Collaborative Computational Project, No. 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763.
- de La Fortelle, E. and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. In *Macromolecular Crystallography*, vol. 276 (Carter, Jr, C. and Sweet, R., eds.), pp. 472–494, Academic Press, New York.
- Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. and Paciorek, W. (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030.
- Perrakis, A., Morris, R. and Lamzin, V. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463.
- Jones, T., Zou, J.-Y., Cowan, S. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119.
- Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132.
- Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881–10890.

Received 18 February 2005/13 April 2005; accepted 19 April 2005

Published as BJ Immediate Publication 19 April 2005, DOI 10.1042/BJ20050310