The AXH module: an independently folded domain common to ataxin-1 and HBP1 ☆

C. de Chiara[a], C. Giannini[a], S. Adinolfi[a], J. de Boer[a], S. Guida[b], A. Ramos[a], C. Jodice[b], D. Kioussis[a], A. Pastore[a,*]

[a]*National Institute for Medical Research, The Ridgeway, London NW71AA, UK*
[b]*Dipartimento di Biologia, Università di Tor Vergata, via della Ricerca Scientifica, 00133 Rome, Italy*

Abstract **Ataxin-1 (ATX1), a human protein responsible for spinocerebellar ataxia type 1 in humans, shares a region of homology, named AXH module, with the apparently unrelated transcription factor HBP1. Here, we describe the first characterisation of the AXH module in terms of its structural properties and stability. By producing protein constructs spanning the AXH modules of ATX1 and HBP1 and by comparing their properties, we have identified the minimal region sufficient for forming independently folded units (domains). Knowledge of the AXH domain boundaries allows us to map many of the interactions of ATX1 with other molecules onto the AXH module. We further show that the AXH of ATX1 is a dimerisation domain and is able to recognise RNA with the same nucleotide preference previously described for the full-length protein. AXH is therefore a novel protein–protein and RNA binding motif.**
© 2003 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Key words: Misfolding disease; Poly-glutamine; Protein module; SCA1; Structure

1. Introduction

Spinocerebellar ataxia type-1 (SCA1), an autosomal-dominant neurodegenerative disorder characterised by ataxia and progressive motor deterioration, is caused by expansion of unstable CAG trinucleotide repeats in the coding region of the corresponding gene [1]. The expansion results in a polyglutamine (polyQ) tract in the gene product, the ataxin-1 (ATX1) protein. Nuclear localisation of expanded ATX1 induces selective death of Purkinje cells in the cerebellar cortex [2]. PolyQ expansion associated with death of specific neuronal cells is the hallmark not only of SCA1, but of a larger family of diseases [3], which share two characteristic features: the pathogenic threshold is ∼35–40 glutamines for most disorders; and the onset is earlier and the progression more severe as the length of polyQ increases. Several hypotheses have been suggested to explain the polyQ role in disease and the threshold between benign and pathological expansions, but the field remains highly controversial. Indications of how to shed light on these issues may come from understanding the cellular role of these proteins in their non-pathological form. So far, ATX1 is one of the best characterised polyQ proteins. It has been reported to have RNA binding activity inversely correlated with the length of its polyQ tract [4]. It is known to self-associate not only through the polyQ but also through a region spanning residues 495–605 [5]. Such multimerisation, however, seems to play a role in progression rather than initiation of the disease [6]. ATX1 has also been implicated in interactions with several other proteins [7–13]. However, despite these findings, the function of ATX1 and the detailed cellular and molecular mechanisms that lead to SCA1 pathogenesis remain obscure.

An efficient way to predict protein function is to search for sequence homology with other proteins [14]. Identification and characterisation of the same motif in otherwise apparently unrelated proteins is therefore considered a powerful tool for detecting conserved units and formulating predictions about their fold and functions.

A small motif (ca 130 amino acids long), named AXH module, was identified in ATX1 and in the apparently unrelated transcription factor HBP1 [15]. HBP1 is thought to be involved in an architectural regulation of chromatin and in specific gene expression [16]. While analysing the sequence of AXH, we noticed that most of the currently known interactions that ATX1 forms with itself and with other proteins comprise the AXH (Fig. 1a), thus suggesting the possibility that this region could fold as an independent domain and form a novel protein–protein and RNA recognition motif.

To test this hypothesis, we initiated a project to study in parallel the two homologous AXH modules. Here, we show that by characterising the state of fold and the stability of the AXH module using a combination of complementary biophysical techniques we can rationalise our current knowledge of the ATX1 functions.

2. Materials and methods

2.1. Sequence analysis

Sequence retrieval and analysis was performed using programmes available from the www.EMBL-heidelberg.de server. Multiple alignments of each of the ATX1 and HBP1 families were obtained independently with the CLUSTALX programme and then merged. Secondary structure was predicted by the PHD programme using the multiple alignment as input.

*Corresponding author. Fax: (44)-20-89064477.
E-mail address: apastor@nimr.mrc.ac.uk (A. Pastore).

☆ Supplementary data associated with this article can be found at doi:10.1016/S0014-5793(03)00818-4

2.2. Protein expression and purification

The construct sequences used in the present study are summarised in Table 1. ATX1 constructs were amplified from human ATX1 (EMBL/GenBank accession no. P54253). HBP1 constructs were produced by PCR amplification of the *Mus musculus* HBP1 sequence (accession no. Q8R316). All constructs were produced as His-tagged GST fusion proteins. The tag was cleaved using Tobacco Etch Virus (TEV) protease cleavage site, which results in the addition of either three (GAM) or four residues (GAMA) at the N-terminus of each construct. Expression was induced in the *Escherichia coli* BL21(DE3) strain using 0.5 mM isopropyl-D-thiogalactopyranoside (IPTG), when the absorbance at 595 nm was 0.5. The first purification step was carried out on a Ni-NTA agarose column (Qiagen). Cleavage of the his-tagged GST was achieved overnight at room temperature using Gibco-BRL His-tagged TEV recombinant protease (cat. no. 10127-017). Separation of GST and TEV from the proteins was performed using a Ni-NTA agarose column. The cleavage product was recovered by elution. The molecular weights of the constructs were checked by electrospray mass spectrometry.

2.3. Analytical ultracentrifugation

Sedimentation equilibrium experiments were carried out using a Beckman XL-A analytical ultracentrifuge equipped with UV absorption optics. Protein concentrations were varied in the range 0.015–0.030 mM in the different experiments but were chosen so that the absorbance at 280 nm was in the 0.2–1 units range (pathlength 12 mm). All the experiments were carried out under reducing conditions (using β-mercaptoethanol), both in 20 mM Tris–HCl buffer (pH 7.0) and 10 mM potassium phosphate, and adding either 50 or 200 mM NaCl. Data were recorded at 16 000, 22 000 and 28 000 rpm. Two datasets were acquired at each speed: the first one 16 h after a speed change, the second one 8 h later to confirm equilibrium conditions and exclude that proteolysis could interfere. The data were analysed with the Origin XL-A/XL1 package (Beckman).
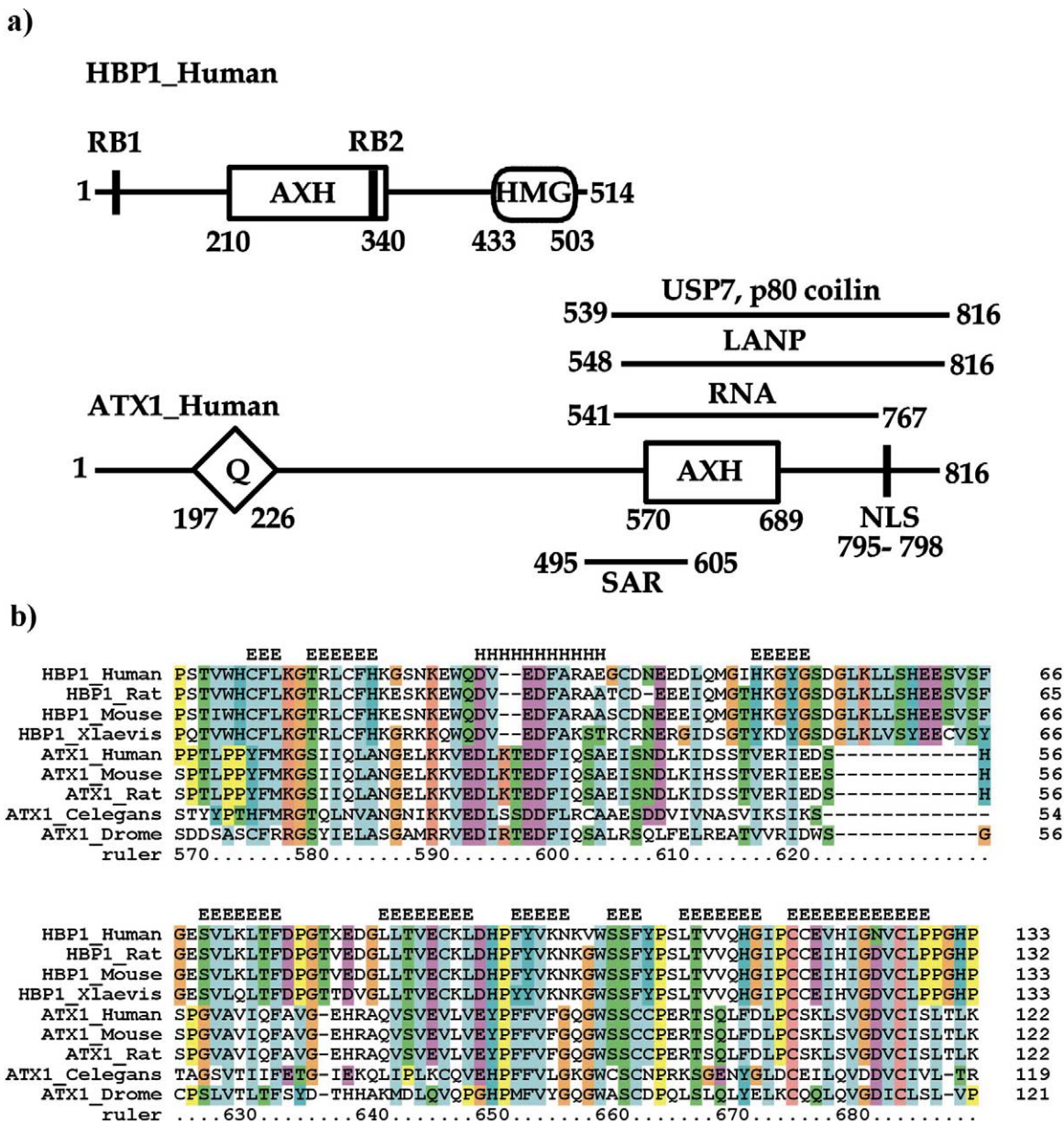

Fig. 1. a: Schematic representation of the domain architecture of ATX1 and HBP1. The positions along the sequence of the high- and low-affinity retinoblastoma binding sites (RB1 and RB2, respectively) and of the HMG and AXH motifs are indicated for HBP1. The polyQ tract (Q), the nuclear localisation signal (NLS) and the AXH module are marked for ATX1. The minimal regions of ATX1 involved in self-association (SAR) [5], RNA [4] and binding to LANP, USP7 and p80 coilin [8,10,11] are also reported. b: Multiple alignment of the AXH sequences of ATX1 and HBP1 families and two hypothetical proteins from *Caenorhabditis elegans* and *Drosophila melanogaster*. Secondary structure prediction for the multiple alignment as obtained by the PHD programme is reported in the first line.

2.4. Circular dichroism (CD) and nuclear magnetic resonance (NMR) measurements

CD measurements were performed on a Jasco J-710 spectropolarimeter equipped with a PTC-843 Peltier system for temperature control. A rectangular quartz thermostated cuvette (Hellma) with 0.1 cm pathlength was used. The spectra were baseline-corrected and the observed ellipticity converted to mean residue molar ellipticity [θ] (deg cm^2 dmol^{-1}). Protein concentration was determined by absorbance using a calculated extinction coefficient at 280 nm and assuming all cysteines to be in a reduced state. The protein concentration typically ranged from 6 to 20 µM in 10 mM phosphate buffer (pH 7). The thermal denaturation curves were obtained following the ellipticity at 222 nm in the range 10–80°C using a heating rate of 1°C/min. The effect of salt concentration on thermal stability was checked by repeating the experiments with no added salt, 50 and 200 mM NaCl.

NMR spectra were recorded at 30°C on a Varian Unity Plus-500 spectrometer operating at 500 MHz ^{1}H frequency. The protein concentration was typically 0.2–0.5 mM in 20 mM Tris–HCl at pH 7, 50 mM NaCl and 2 mM β-mercaptoethanol.

2.5. RNA-binding assay

RNA binding to homopolymers was probed by following the assay described by Yue et al. [4]. 0.5 ml of 3 µg solutions of the purified uncleaved His-tagged GST-fused proteins in the binding buffer (10 mM Tris–HCl at pH 7.4, 2.5 mM MgCl$_2$, 0.5% Triton, 150 mM NaCl) were individually mixed with agarose-bound poly(rA), poly-(rU), poly(rG) and poly(rC) beads (Sigma or Amersham) (50 µl of fully swollen matrix in the binding buffer). The mixture was left to incubate for 4 h at 4°C. The beads were pelleted with a short spin in a microfuge and washed four times with 600 µl binding buffer prior to resuspension in SDS–PAGE loading buffer. Bound proteins were eluted from nucleic acid by boiling and redissolved in an SDS–polyacrylamide gel. The proteins were transferred to PDVF membrane (Amersham) and visualised by immunoblot. Filters were processed and probed using a monoclonal anti-6-His-tag antibody. The assay was repeated twice to check reproducibility.

3. Results

3.1. HBP1 and ATX1 share a common motif within a different architecture

HBP1 and ATX1 have a different molecular architecture and the two AXH domains are differently located along the primary sequence (Fig. 1a). In ATX1, AXH spans residues 570–689 (the numbering is referred to human ATX1), quite distant both from the polyQ tract (residues 197–226) and the nuclear localisation signal (NLS) (residues 795–798) [2]. AXH partially overlaps the self-association region (SAR, residues 495–605) [5] and is contained in the region involved in RNA binding (residues 541–767) [4]. In HBP1, the AXH domain is located approximately in the middle of the protein and is followed by a C-terminal high-mobility group (HMG), a domain known to bind DNA in a sequence-dependent fashion [17]. Of the two retinoblastoma binding (RB) sites identified in HBP1 [18], the-low affinity one, which consists of an IPCCE motif, is at the C-terminus of the AXH module. The region 192–400, containing AXH, is believed to be the minimal HBP1 region essential for the interaction with the T-cell factor proteins [19].

The AXH module is highly conserved within each of the two protein families, sharing ca 78–80% identity, having ∼28% identity and ∼54% similarity between the ATX1 and HBP1 families (Fig. 1b). An AXH module is also present in two shorter hypothetical proteins from *D. melanogaster* and from *C. elegans*. Their sequences cluster with the AXH of ATX1. Two cysteines are completely conserved throughout both families, suggesting that these residues have a functional or structural role. Although relatively low, the degree of similarity is uniformly distributed along the module sequence, with only one long insertion of 10 residues in HBP1 and no low-complexity sequence regions. This makes us confident that AXH must share the same three-dimensional fold in both protein families. Secondary structure prediction on a multiple sequence alignment predicts a content of 7% helix, 44% β and 49% loop structure.

3.2. The AXHs from HBP1 and ATX1 fold into autonomous domains

Domain definition does not always coincide with a sequence motif. To determine the minimal length to obtain a folded and well-behaved AXH from both ATX1 and HBP1, constructs for each of the two proteins were initially designed (ATX1(568–689) and HBP1(208–345)) that encompass the region of maximal homology (Table 1). Both proteins were expressed in good yields.

The secondary and tertiary structures were checked by CD and NMR, respectively. The far-UV CD spectra of the constructs have similar features (Fig. 2a), with a single minimum around 201 nm and no positive bands between 190 and 200 nm. This behaviour is typical of proteins with a predominance of β-structure and loop regions. Thermal denaturation monitored by CD showed that HBP1(208–345) unfolds with a co-operative biphasic reversible transition with midpoints at 49.4 and 59.8°C, indicating melting of different parts of the molecule with different stabilities (see supplementary material). Unfolding of ATX1(568–689) is reversible but non-cooperative, possibly suggesting that the domain stability is so high that no transition can be detected in the temperature range explored. Increasing salt concentrations resulted in a minor (up to 3°C) destabilisation effect for HBP1(208–345) and no detectable change for ATX1(568–689).

Table 1
Summary of the constructs used in the present study and of their properties

Constructs	Boundaries	Behaviour	T_m (°C)		
			0 mM NaCl	50 mM NaCl	200 mM NaCl
HBP1(208–335)	PSTVWHCFLK…VCL	Inclusion bodies	n.d.[a]	n.d.[a]	n.d.[a]
HBP1(208–345)	PSTVWHCFLK…VCLPPGHPDAINF	Folded, monomer	49.4/59.8	48.6/58.8	46.2/59.0
HBP1(200–343)	DLGWCNSWPSTVWHCFLK…VCLPPGHPDAI	Folded, monomer	49.8/57.2	48.0/56.6	49.6/56.8
ATX1(556–700)	PVVQSVASPAAAPPTLPPYFMK…VCISLTLKNLKNGSVKKGQ	Aggregated	52.8	51.8	51.6
ATX1(568–689)	PPTLPPYFMK…VCISLTLK	Folded, dimer	n.c.[b]	n.c.[b]	n.c.[b]
ATX1(574–689)	YFMK…VCISLTLK	Low yield	n.d.[a]	n.d.[a]	n.d.[a]

The melting points (T_m) as recorded by CD at increasing ionic strengths are also reported.
[a]Not determined.
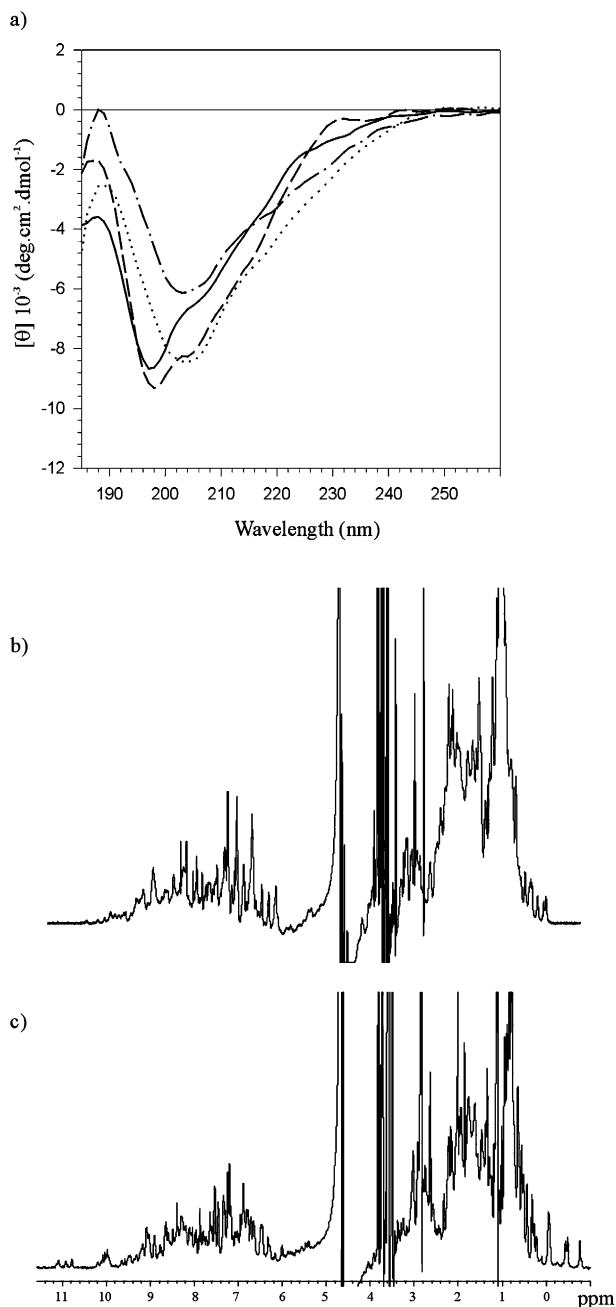[b]Non-cooperative transition.

a)

b)

c)

Fig. 2. a: Far-UV CD spectra of HBP1(208–345) (continuous line), HBP1(200–343) (dashed line), ATX(568–689) (chained line) and ATX(556–700) (dotted line) recorded at 20°C. 1D ^{1}H NMR spectra of ATX1(568–689) (panel b) and HBP1(208–345) (panel c).

1D ^{1}H NMR spectra confirm that the two constructs are stably folded: the resonances are well-dispersed, with relatively narrow linewidths (Fig. 2b). The presence of up-field resonances in the range −1 to 0.8 ppm arising from aliphatic protons in persistent proximity to aromatic groups is also typical of folded proteins.

3.3. The domain boundaries of AXH are clearly defined

To check whether the AXH module could be shortened further, the constructs ATX1(574–689) and HBP1(208–335) were made, omitting an N-terminal and a C-terminal pro-

line-rich stretch, respectively (Table 1). The rationale for this choice was that neither stretch is conserved and their sequences do not suggest secondary structure elements. ATX1(574–689) was expressed in the soluble fraction but the yields were extremely low. Both the CD and 1D NMR spectra of the construct showed that the fragment retains some 3D fold but adopts a conformation different from that of the longer version (see supplementary material). Most of the expressed HBP1(208–335) formed inclusion bodies despite all attempts to shift the equilibrium towards the soluble form.
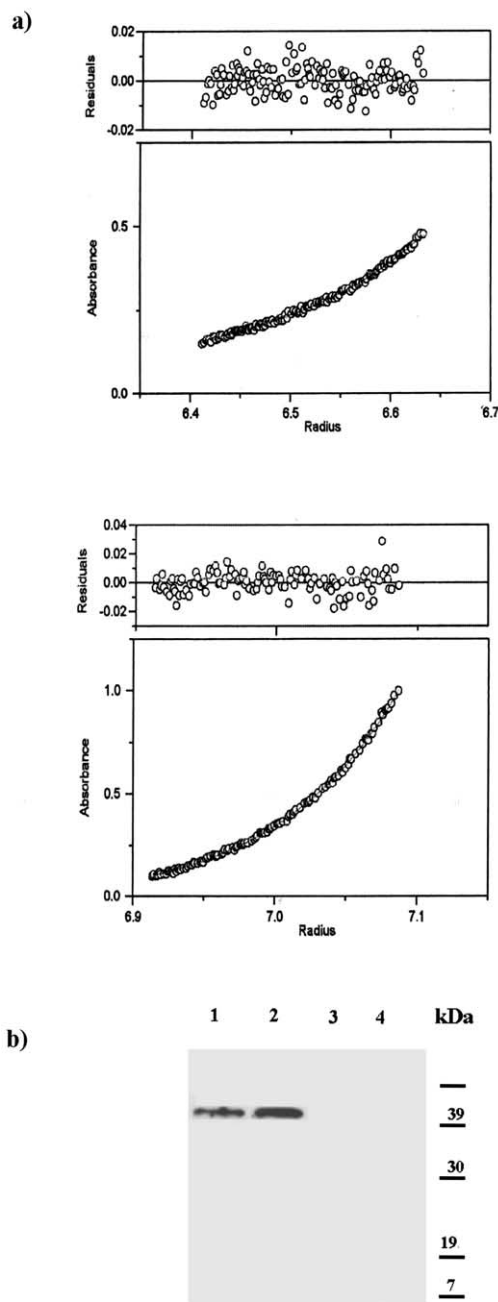
a)

b)

Fig. 3. a: Analytical equilibrium ultracentrifugation data of ATX1(568–689) (top) and HBP1(208–345) (bottom). The data were recorded at 15°C and 22 000 and 28 000 rpm, respectively. b: RNA-binding assays using homopolymers p(rG), p(rU), p(rA), p(rC) immobilised on agarose beads (lanes 1, 2, 3 and 4, respectively). The markers and their molecular weights are indicated in the last lane.

To investigate a possible role of the flanking residues in domain stability, one more construct was designed for each of the two proteins by extending the boundaries both N- and C-terminally. In the absence of other indications, we selected the new boundaries to include predicted secondary structure elements. The thermal stabilities of the constructs were compared following the rationale that, if the amino acid extensions interact with AXH, different melting points should be observed for the extended proteins as compared to the non-extended versions. The HBP1(200–343) construct remains reversible and biphasic, with the two unfolding midpoints being comparable to those of the non-extended version (49.8 and 57.2°C). Denaturation of the extended ATX1(556–700) is cooperative, with the midpoint at 52.8°C, but the protein precipitates at high temperature. Accordingly, the NMR spectra of HBP1(200–343) and ATX1(556–700) are those of folded species and reminiscent of those of HBP1(208–345) and ATX1(568–689). However, the quality of the spectrum of ATX1(556–700) is definitely inferior, with a strong line broadening that suggests formation of large aggregates (data not shown).

These results show that the stability of AXH is mostly independent of flanking regions but depends on the domain integrity: deletion of even a few residues leads to unfolding proteins.

3.4. The AXH module of ATX1 but not of HBP1 forms dimers

The ATX1 and HBP1 constructs were probed for their association state by analytical ultracentrifugation (Fig. 3a). The apparent molecular weight of 18.0 kDa derived for HBP1(208–345) approximates within experimental error the monomer value (16 410 Da). Sedimentation equilibrium curves of ATX1(568–689) can be fit either as a dimer (at 200 mM NaCl) or as a mixture of a prevalent dimeric species in equilibrium with further aggregates (at 50 mM NaCl). A higher tendency to aggregate was observed for the extended ATX1(556–700) which, depending on salt conditions, is either a tetramer or another multimer in solution. This is consistent with the broadening observed in the NMR spectrum of this construct and confirms that extension of AXH of ATX1 promotes further aggregation. In contrast, HBP1(200–343) remains monomeric.

These results show that AXH is in a different aggregative state in the two protein families: while the AXH of HBP1 forms a well-behaved monomer, the AXH of ATX1 exists in solution as a dimer.

3.5. AXH contains an RNA binding motif

RNA binding was tested using the four homopolymers immobilised on agarose beads as described in [4]. ATX1(568–689) is able to bind the homopolymers when 150 mM NaCl concentration was used (Fig. 3b). Poly(rG) and poly(rU) are recognised with approximately the same affinities, whereas no binding is observed for poly(rA) and poly(rC).

4. Discussion

Besides giving the first characterisation of a novel domain in terms of its structural properties and stability, the work presented here provides the first direct structural information about ATX1. We have shown that deletion mutants that cut out even a few residues within the domain boundaries defined by our results lead to unfolded proteins. This observation has direct implications for interpreting mapping studies of the interaction sites of ATX1 with LANP, USP1 and p80 coilin [8,10,12]. Only constructs which comprise the full-length AXH module are able to retain binding, whereas truncation of AXH abolishes binding as expected if its tertiary structure was disrupted. Analysis of the published deletion mutants also strongly suggests that AXH of ATX1 would be able to bind to RNA [4]: RNA binding has been demonstrated for full-length ATX1 and for two fragments containing residues 277–816 and 1–757. Conversely, mutants in which AXH is excluded (i.e. 1–540, 1–489 and 1–425) completely lose RNA binding capability. Accordingly, we have demonstrated that the AXH of ATX1 is able to bind to RNA homopolymers with the same sequence preference observed for the full-length protein [4]. The observation that the efficiency of RNA binding inversely correlates with the length of the polyQ tract suggests that the tertiary arrangement of ATX1 could bring close in space the AXH and the N-terminus. Expansion of the polyQ tract could either mask the RNA binding sites in AXH or promote aggregation, which in turn would reduce accessibility to RNA.

Even more interesting in light of the SCA1 pathology, thought to be linked to protein misfolding and aggregation, are our results about the self-association and the unfolding properties of AXH. Suggestion that AXH could be responsible for ATX1 self-association comes both from noticing that the SAR overlaps with AXH and from loss of association observed for deletion mutants that truncate the N-terminus of the AXH module [5]. Here, we prove conclusively that the AXH of ATX1 forms dimeric species, thus showing that the observed self-association is a specific dimerisation of the AXH module. While providing a possible explanation for an unusual stability against unfolding of ATX1, dimer formation is, however, not an intrinsic feature of AXH, since the AXH of HBP1 is monomeric.

In conclusion, we have used the sequence similarity between regions of two otherwise uncorrelated proteins to obtain information about their domain architecture. Definition of a structural domain in ATX1 has allowed us to interpret the functional data available and to understand whether a deletion mutant loses its function because of direct loss of a functional element, or of a region essential for the three-dimensional scaffold. Further insights into the role of AXH may be provided by three-dimensional structure determination of AXH, which is currently in progress in our laboratory. While awaiting this information, we must conclude that the AXH motif is a novel protein–protein and RNA-binding domain that seems to play an important role in the functions of ATX1.

References

[1] Zoghbi, H.Y. and Orr, H.T. (1995) Semin. Cell Biol. 6, 29–35.
[2] Klement, I.A., Skinner, P.J., Kaytor, M.D., Yi, H., Hersch, S.M., Clark, H.B., Zoghbi, H.Y. and Orr, H.T. (1998) Cell 95, 41–53.
[3] Masino, L. and Pastore, A. (2002) Brain Res. Bull. 56, 183–189.

[4] Yue, S., Serra, H.G., Zoghbi, H.Y. and Orr, H.T. (2001) Hum. Mol. Genet. 10, 25–30.

[5] Burright, E.N., Davidson, J.D., Duvick, L.A., Koshy, B., Zoghbi, H.Y. and Orr, H.T. (1997) Hum. Mol. Genet. 6, 13–18.

[6] Skinner, P.J., Vierra-Green, C.A., Emamian, E., Zoghbi, H.Y. and Orr, H.T. (2002) Neuromol. Med. 1, 33–42.

[7] Koshy, B., Matilla, T., Burright, E.N., Merry, D.E., Fischbeck, K.H., Orr, H.T. and Zoghbi, H.Y. (1996) Hum. Mol. Genet. 5, 1311–1318.

[8] Matilla, A., Koshy, B.T., Cummings, C.J., Isobe, T., Orr, H.T. and Zoghbi, H.Y. (1997) Nature 389, 974–978.

[9] Davidson, J.D., Riley, B., Burright, E.N., Duvick, L.A., Zoghbi, H.Y. and Orr, H.T. (2000) Hum. Mol. Genet. 9, 2305–2312.

[10] Hong, S., Kim, S.J., Ka, S., Choi, I. and Kang, S. (2002) Mol. Cell. Neurosci. 20, 298–306.

[11] Okazawa, H., Rich, T., Chang, A., Lin, X., Waragai, M., Kajikawa, M., Enokido, Y., Komuro, A., Kato, S., Shibata, M., Hatanaka, H., Mouradian, M.M., Sudol, M. and Kanazawa, L. (2002) Neuron 34, 701–713.

[12] Hong, S., Ka, S., Kim, S., Park, Y. and Kang, S. (2003) Biochem. Biophys. Acta 1638, 35–42.

[13] Chen, H.K., Fernandez-Funez, P., Acevedo, S.F., Lam, Y.C., Kaytor, M.D., Fernandez, M.H., Aitken, A., Skoulakis, E.M., Orr, H.T., Botas, J. and Zoghbi, H.Y. (2003) Cell 113, 457–468.

[14] Bork, P. and Gibson, T. (1996) Methods Enzymol. 266, 162–184.

[15] Mushegian, A.R., Bassett Jr., D.E., Boguski, M.S., Bork, P. and Koonin, E.V. (1997) Proc. Natl. Acad. Sci. USA 94, 5831–5836.

[16] Lesage, F., Hugnot, J.P., Amri, E.Z., Grimaldi, P., Barhanin, J. and Lazdunski, M. (1994) Nucleic Acids Res. 22, 3685–3688.

[17] Zhuma, T., Tyrrell, R., Sekkali, B., Skavdis, G., Saveliev, A., Tolaini, M., Roderick, K., Norton, T., Smerdon, S., Sedgwick, S., Festenstein, R. and Kioussis, D. (1999) EMBO J. 18, 6396–6406.

[18] Lavender, P., Vandel, L., Bannister, A.J. and Kouzarides, T. (1997) Oncogene 14, 2721–2728.

[19] Sampson, E.M., Haque, Z.K., Ku, M.-C., Tevosian, S.G., Albanese, C., Pestell, R.G., Eric Paulson, K. and Yee, A.S. (2001) EMBO J. 20, 4500–4511.