The Structure of the AXH Domain of Spinocerebellar Ataxin-1*

Yu Wai Chen‡§, Mark D. Allen‡, Dmitry B. Veprintsev‡, Jan Löwe‡, and Mark Bycroft¶

From the ‡Centre for Protein Engineering and the §Laboratory of Molecular Biology, Medical Research Council Centre, Hills Road, Cambridge CB2 2QH, United Kingdom

Spinocerebellar ataxia type 1 is a late-onset neurodegenerative disease caused by the expansion of a CAG triplet repeat in the SCA1 gene. This results in the lengthening of a polyglutamine tract in the gene product ataxin-1. This produces a toxic gain of function that results in specific neuronal death. A region in ataxin-1, the AXH domain, exhibits significant sequence similarity to the transcription factor HBP1. This region of the protein has been implicated in RNA binding and self-association. We have determined the crystal structure of the AXH domain of ataxin-1. The AXH domain is dimeric and contains an OB-fold, a structural motif found in many oligonucleotide-binding proteins, supporting its proposed role in RNA binding. By structure comparison with other proteins that contain an OB-fold, a putative RNA-binding site has been identified. We also identified a cluster of charged surface residues that are well conserved among AXH domains. These residues may constitute a second ligand-binding surface, suggesting that all AXH domains interact with a common yet unidentified partner.

Spinocerebellar ataxia type 1 (SCA1) is an autosomal dominat neurodegenerative disorder characterized by the loss of Purkinje’s cells in the cerebellar cortex. It is a member of the polyglutamine expansion disease family. These diseases are caused by the abnormal lengthening of a CAG triplet repeat in the coding region of the respective gene. In normal individuals, the triplet repeat translates into a polymorphic glutamine tract of fewer than 35–40 residues. In patients with these diseases, the length of the polyglutamine tract exceeds the 35–40-residue threshold. This results in a toxic gain of function that leads to tissue-specific neuronal loss. In most cases, nuclear ubiquitinated aggregates of the pathogenic proteins are observed in affected neurons (1). The relationship between the polyglutamine tract and disease pathology is unclear and has been the subject of much interest (2). In some cases, overexpression of even the normal protein can lead to mild disease phenotypes (3), suggesting that protein misfolding or turnover may play a role in the disease process.

SCA1 is caused by polyglutamine expansion in ataxin-1, a nuclear protein of ~800 residues. Transgenic animal models for this disease have contributed significantly to our understanding of polyglutamine expansion diseases in general. The areas addressed have included the role of protein aggregates in neural toxicity, the effects of chaperones and the proteasome in neuropathology, and the role of post-translational modification and protein-protein interactions in disease progression (4). Recently, it has been shown that neurodegeneration is mediated by the interaction of ataxin-1 with the 14-3-3 proteins (5). This finding has emphasized the importance of understanding the normal function of the protein. Ataxin-1 appears to be involved in the regulation of gene expression. It has been shown to associate with several proteins involved in controlling transcription, and an expanded allele of Sca1 can down-regulate early gene expression in Purkinje’s cells in transgenic mice (6). A 120-residue region of ataxin-1 is similar in sequence to part of the HMG box transcription factor HBP1 (HMG box-containing protein-1) (Fig. 1) (7–10). This protein has a role in chromatin remodeling and regulates gene expression during the arrest of cell proliferation and during cell differentiation. The region of similarity has been shown to act as a transcription repression domain in HBP1 (9). This region has been termed the AXH (ataxin-1/HBP1) domain (SMART Database accession number SM00536). There are small proteins corresponding to just the AXH domain present in Caenorhabditis elegans and Drosophila melanogaster (Fig. 1), suggesting that this is an independently folded unit. The ataxin-1 AXH domain has been implicated in self-association and RNA binding (11, 12) as well as in binding of a ubiquitin protease (13) and p80 coilin (14). As part of an effort to understand the structure and function of proteins involved in polyglutamine expansion diseases, we have determined the structure of the AXH domain of ataxin-1.

EXPERIMENTAL PROCEDURES

Crystallization and Data Collection—The DNA sequence encoding the ataxin-1 AXH domain (residues 563–694, all residue numbering in this work follows that of ataxin-1 with a 30-glutamine repeat, unless otherwise stated) was subcloned into a modified pRSET-A vector (Invitrogen) that allows overproduction of a C-terminal His-tagged protein with a thrombin cleavage site. The plasmid was transformed into the Escherichia coli C41 strain (15) for overexpression. A double mutant of the AXH domain containing the I611M and V641M mutations was created to facilitate selenium incorporation. This mutant, SeMet AXH, was overproduced using the protocol of van den Ent et al. (16). Both native and SeMet AXH proteins were purified by Ni2+ affinity chromatography, subjected to thrombin cleavage, and further purified by anion-exchange chromatography (Amersham Biosciences Resource-Q) and gel filtration (Superdex-75). The purified recombinant AXH proteins have 133 residues and contain a non-native Gly residue at their N termini. The identities of the proteins were confirmed by electrospray mass spectrometry.

The proteins were concentrated to ~20 mg/ml in 10 mM Tris (pH 7.0) with 5 mM dithiothreitol. Both native and SeMet AXH crystals were

Received for publication, September 4, 2003, and in revised form, October 23, 2003
Published, JBC Papers in Press, October 28, 2003, DOI 10.1074/jbc.M309817200

© 2004 by The American Society for Biochemistry and Molecular Biology, Inc.
Printed in U.S.A.
Structure of the AXH Domain of Ataxin-1

The values in parentheses are for the highest resolution shell (1.79 to 1.70 Å for the native set and 3.16 to 3.00 Å for the SeMet sets).

| TABLE I | Data collection and processing statistics |
|---------|------------------------------------------|
| Native  | SeMet                                    |
| Wavelegth (Å) | 0.960 | 0.97916 | 0.97966 | 0.960 |
| Resolution (Å) | 7.0 | 7.0 | 7.1 | 7.0 |
| Observed reflections | 267,655 | 87,808 | 88,850 | 88,104 |
| Unique reflections | 62,044 | 11,843 | 11,858 | 11,841 |
| Multiplicity | 4.3 | 4.3 | 4.3 | 4.3 |
| Completeness (%) | 99.6 | 99.8 | 99.9 | 99.9 |
| Rmerge (%) | 5.9 | 8.8 | 9.0 | 6.0 |
| Rmerge (%) | 6.5 | 4.1 | 7.5 | 11.8 |
| TABLE II | Multiple wavelength anomalous dispersion phasing statistics |
| The values reported here are for all data to 3.0 Å. |
| Correlations | Peak vs. edge | 0.38 | Peak vs. remote | 0.43 | Edge vs. remote | 0.25 |
| Mean figure of merit | After SOLVE | 0.43 | After SHARP | 0.48 |
| TABLE III | Crystallographic data and refinement statistics for the ataxin-1 AXH domain |
| Resolutions range (Å) | 70.7–1.70 | 70.7–1.70 |
| Rmerge (%) | 21.0 | 25.2 |
| Average B factors (Å²) | 25.8 | 25.1 |
| r.m.s.d. deviation from ideality | 33.5 |
| Bond angles | 0.020 |
| Coordinate error (Å) | 1.78 |
| Geometry: Ramachandran plot analysis |
| In most favored region (%) | 99.1 |
| In additionally allowed region (%) | 6.7 |
| In generously allowed/disallowed regions (%) | 0.2/0.0 |

Sequence Conservation—Five AXH domains (human ataxin-1, CG4547 protein from D. melanogaster)
Q9W3V7), K04F10.1 protein from C. elegans (Swiss-Prot accession number O44771), human HBP1 (Swiss-Prot accession number O60381), and frog HBP1 (Swiss-Prot accession number Q8JH80) were submitted to the ConSurf server for analysis. The ConSurf server assigns relative conservation scores to each residue, taking into account the evolutionary relationships among the family of homologs. The scores are normalized such that the average score is zero, and negative and positive deviations represent the degrees of conservation and variation, respectively. Each residue is then assigned a value 1–9 (1 for most variable, 5 for average, up to 9 for most conserved), which is used for mapping the relative conservation on the molecular surface (see figure legends).

**Dimeric Interface Analysis**—The dimeric interfaces were analyzed with the Protein-Protein Interaction server at University College London.

**Analytical Ultracentrifugation**—Analytical centrifugation experiments were performed with a Beckman XL-I ultracentrifuge in 50 mM Tris (pH 7.5) and 100 mM NaCl at 293 K. Protein sample loading concentrations were 0.5, 1, and 10 mg/ml. Absorbance at 280 nm was measured.
FIG. 5. Interactions at the AXH C-D dimeric interface. A, chain D is shown with secondary structures in orange, and the N-terminal tail of 21 residues (positions 562–582) on chain C is in blue. This view is the same as that in Fig. 4B. B, shown is a detailed view of residues 562–582 of chain C (blue) and chain D (orange). The dyad axis is shown in yellow. Note that the two chains differ in structure up to residue 579.

**TABLE V**

Most important residues for dimer interaction

The residues with ΔASA > 60 Å² upon dimerization are tabulated. The conservation scores were calculated with the ConSurf server (see “Experimental Procedures†), MC and SC, main chain and side chain contributions, respectively.

|                 | Chain C | Chain D | Sequence conservation score |
|-----------------|---------|---------|----------------------------|
| Ala566          | 67 (MC + SC) | -0.28 (conserved) |
| Pro568          | 93 (SC) | 109 (MC + SC) | -0.35 (conserved) |
| Pro569          | 113 (MC + SC) | 109 (MC + SC) | 3.57 (highly variable) |
| Thr570          | 93 (SC) | -0.22 (conserved) |
| Tyr574          | 80 (SC) | 77 (SC) | -0.25 (conserved) |
| Phe575          | 162 (MC + SC) | -0.64 (highly conserved) |
| Met576          | 63 (SC) | -0.41 (conserved) |
| His580          | 70 (MC + SC) | 65 (MC + SC) | 0.53 (highly variable) |
| Gln582          | 70 (MC + SC) | 72 (MC + SC) | 0.14 (variable) |
| Leu589          | 78 (SC) | -0.09 (variable) |
| Leu595          | 94 (SC) | 113 (SC) | -0.02 (variable) |

measured at 12,000, 18,000, and 24,000 rpm using interference optics and analyzed using UltraSpin software.

**RESULTS**

**Quaternary Structure**—There are four chains in the crystal asymmetric unit arranged into two spherical dimers (A-B and C-D) in which individual molecules are related by an 2-fold rotation (Fig. 2). Analytical ultracentrifugation indicated that the AXH domain is a dimer in solution (Fig. 3). The A-B and C-D dimers form similar interfaces with buried surface areas of 1570 and 1645 Å², respectively. These values agree well with those obtained for other known homodimeric interfaces (Table IV) (31). The other intermolecular contacts (B-C) seen in the crystal lattice are most likely an effect of crystal packing.

**Monomer Structure**—The main feature of the monomer is an open β-barrel with a Greek key motif formed by strands β3, β4, β5, and β9 (Fig. 4). This is known as an OB (oligomer-binding)-fold and is found in many different proteins (32–34). When the AXH domain was queried with the DALI server6 for structural homologs, the four top scoring structures were all OB-fold proteins, the most similar being the N-terminal domain of ribosomal protein L2. The AXH domain lacks the strand corresponding to OB-fold strand 5. In a typical OB-fold structure, the connectivity between strands 3 and 4 forms an α-helix that caps the barrel. In the AXH domain, the loop between strands β5 and β9 also contains a helix. There are, however, insertions at both ends of the loop that form three short strands (β6, β7, and β8) (Fig. 4). The OB-fold is preceded by an α-helix (α1) that packs onto the edge of the β-barrel and two short strands (β1 and β2) (Figs. 4B and 6B).

**Differences among the Four Monomers**—There are considerable structural differences among the four chains of the AXH domain in the crystal asymmetric unit. Most noticeably, the 19 N-terminal residues preceding strand β1 (residues 562–580) can have two different structures: the A and C monomers adopt a main chain conformation that is distinct from that of the B and D monomers (Fig. 5, A and B). In chains A and C, residues 565 and 566 constitute a strand (β0), and residues 573–575 constitute a partial 310 helix (h1) (Fig. 5A). In chains B and D, strand β0 is replaced by a 310 helix (h0) formed by residues 568–570 (chain B) or by residues 564–566 (chain D), followed by residues 574–576 constituting the second 310 helix (h1) (Fig. 5A). In a strict structural sense, A-B and C-D are heterodimers, with residues 581–688 being the invariant body of the protein. When chain A is aligned with chains B–D over this range, the

---

*Available at www.mrc-cpe.cam.ac.uk.

†Available at www.ebi.ac.uk/dali.
**Fig. 6. Sequence conservation of the ataxin-1 AXH domain.** A, view of the C-D dimer (chain C in blue and chain D in orange) showing charged surface residues that are highly conserved among the AXH domains. These residues may contribute a conserved binding surface of a yet unidentified interaction partner. The RNA-binding site is inferred from the general ligand-binding surface of other OB-fold proteins (see “Discussion”). B, sequence alignment of five AXH domains: human ataxin-1, CG4547 protein from D. melanogaster (Swiss-Prot accession number Q9W3V7), K04F10.1 protein from C. elegans (Swiss-Prot accession number O44771), human (h) HBP1 (Swiss-Prot accession number O60381), and...
and that dimerization is an important requirement for its normal function. The structure of the AXH domain and the analytical ultracentrifugation data provide direct confirmation of this observation. From truncation studies, residues 495–605 have been identified as the minimal self-association region of ataxin-1 (Fig. 1) (12). The boundaries of the self-association region do not coincide with those of the AXH domain: the two regions share ~40 residues in common, including the 20 residues that are the determinants for dimerization (Fig. 1). The fact that the self-association region can self-associate indicates that the interactions between the two N-terminal tails are strong enough to mediate dimerization in vitro, even without the rest of the AXH domain. A recent study revealed that the first five residues (positions 563–567, SPAAA) can be removed without affecting dimerization, whereas a further deletion of residues 568–573 (PPTLPP) leads to low yield and non-native structure (36).

It is conceivable that these dimerization residues can be exploited in mediating heterodimeric interactions with other proteins. When the 20-residue peptide sequence was queried in BLASTP for similarity, the only significant hits were from the equivalent regions in the mouse and rat ataxin-1 homologs (data not shown). The search did not even pick up the corresponding region in the HBP1 protein, presumably because the overall homology of this 20-residue motif is weak (Fig. 6B). Nevertheless, the best conserved residues (Lys577, Gly578, and Ser579) in this 20-residue motif contribute little to dimerization (Fig. 6B). It is unlikely that the AXH domain and HBP1 interact via this 20-residue motif.

The 20 N-terminal residues constitute an interesting motif that can adopt one of two structures, and these two different conformations interact complementarily to form an extensive dimeric interface. Despite the low sequence complexity, these two alternative conformations are well ordered (Fig. 5B). Apparently this structural adaptability is essential in maintaining the AXH dimer. It has been known for some time that some protein or peptide sequences can adopt different conformations dependent on the protein context. The most dramatic demonstration is a designer protein that harbors two 11-residue motifs, called the “chameleon sequence,” that can form different secondary structures on different parts of the protein (37). There are many cases of disordered loops in proteins becoming well ordered upon binding a ligand. When a dimeric protein binds to a single DNA or RNA substrate, part of the monomer can adopt variable structures to allow for asymmetric interactions. This is found in the MutS/MCM1-DNA complex (38), the MutS-DNA complex (39), and the NSP3-mRNA complex (40). The 20-residue chameleon sequence in the AXH domain has several unique features compared with these structures. First, the existence of alternative structures is not induced by ligand binding. On the contrary, the two unique conformations can be viewed as mutually inducing or adapting. Second, the interactions contributed by this motif are extensive and involve 20 residues compared with other reported chameleon sequences, which usually involve fewer than 11 residues. To the best of our knowledge, this observation of a dimeric interface...
Structure of the AXH Domain of Ataxin-1

Ataxin-1 is known to be able to bind to RNA (but not to DNA) via the region defined by residues 541–767, including the whole AXH domain (Fig. 1) (11). Importantly, the binding was found to be affected by the length of the glutamine repeat. Sequence analysis failed to reveal any homology between this region and other known RNA-binding motifs (11). A more recent study demonstrated that, within this region, it is the AXH domain (residues 568–689) that binds RNA homopolymers poly(rG) and poly(rU) (36). Here, we have shown that the AXH domain has an OB-fold. Many proteins with this topology are known to bind nucleic acids and, in particular, RNA. Ribosomal proteins S1 and L2, the Rho transcription termination factor, and translation initiation factor eIF1A all harbor RNA-binding OB-fold domains (41–44). OB-fold proteins share a general ligand-binding surface (32, 33). When projected on the AXH domain, the corresponding OB-fold general binding site comprises strands β3, β4, and β5 and the loops between them. This site is not obstructed by dimerization and contains several positively charged and aromatic residues that are typically involved in protein-RNA interactions (Fig. 6A). These putative RNA-binding residues are not conserved in other AXH domains (Fig. 6B). It is not known if RNA binding is a general property of AXH domains.

There is a cluster of charged residues that are highly conserved among the AXH domains (Fig. 6, A and B) that could constitute a ligand-binding site. If this were the RNA-binding site, it would represent a novel form of nucleic acid recognition for an OB-fold.

The C-terminal region of ataxin-1, consisting of residues 539–816, including the whole AXH domain, is responsible for interaction with a ubiquitin-specific protease, USP7 (Fig. 1) (13). This interaction is, again, influenced by the length of the polyglutamine tract of ataxin-1. It has also been found that USP7 interaction is disrupted when the crucial dimerization residues (positions 563–582) of the AXH domain are removed. This suggests that either these residues are involved in binding USP7 or that the AXH domain interacts with USP7 as a dimer.

Implications for the Function of HBP1—HBP1 is a sequence-specific DNA-binding protein that can act as both a transcription activator and repressor. HBP1 is believed to function by remodeling chromatin. The three-dimensional structure of HBP1 is unknown, except for the small three-helix DNA-binding HMG box near the C terminus. Based on the structure reported here, it can be assumed that HBP1 also contains an OB-fold module. Recent work has shown that the AXH domain of HBP1 is monomeric and that dimerization is not a general function of AXH domains (36). The AXH domain of HBP1 overlaps with a known transcription repression domain. The putative second ligand-binding site is strongly conserved in the AXH domains of HBP1 and ataxin-1 (Fig. 6B), suggesting that they may bind to similar targets. It is possible that the AXH domain of HBP1 also binds RNA.

HBP1 is a target of the members of the retinoblastoma (RB) protein family (9). The AXH domain of HBP1 contains the sequence LXXCE (Fig. 6B), which is similar to the consensus RB protein-binding sequence, LXXCE. The structures of several proteins complexed to the RB protein have been determined (45, 46). In all of these, the LXXCE motifs are in an extended conformation and are fully exposed. The region equivalent to the LXXCE motif in HBP1 in the ataxin-1 AXH domain shows a high degree of conservation (Fig. 6B), suggesting that the structure of this region will be similar in the two proteins. Based on this, it can be inferred that the LXXCE motif of HBP1 is buried and inaccessible to the RB protein (Fig. 7). It is more likely that RB protein binding occurs at a second site containing the LXXCE motif in the N-terminal portion of HBP1.

Conclusion—The structure of the AXH domain provides the first three-dimensional information on ataxin-1, a polyglutamine expansion disease protein. This has allowed functional data on the protein to be put into a structural context. The structure offers supporting evidence for a RNA-binding function for ataxin-1. Ataxin-1, another member of the polyglutamine expansion disease family, also has a region of sequence similarity to proteins involved in RNA splicing (47) and associates with an RNA-binding protein (48). This suggests that disruption of protein-RNA interactions might be a common feature of these diseases. There is considerable evidence that ataxin-1 is involved in the control of gene expression. Noncoding RNA molecules have been shown to play a role in transcription regulation (49), and it is possible that the AXH domain of ataxin-1 binds to a regulatory RNA. Given the role of HBP1 in transcription regulation, it is possible that the AXH domain of ataxin-1 has a similar role. It will be important to identify the physiological ligand(s) for these domains, as this will help to delimit the role of both ataxin-1 and HBP1 in the regulation of gene expression.

What does this structure tell us about the likely effects of polyglutamine expansion in SCA1 on ataxin-1 function and stability? The ataxin-1 AXH domain is remote from the glutamine repeat in sequence (Fig. 1). The length of the polyglutamine tract does not affect dimerization (12), presumably because the AXH dimer has high stability. The dimer has extensive complementary interfaces and is consistent with noncooperative thermal unfolding behavior (36). Interactions of ataxin-1 with RNA and USP7 are weakened as the glutamine repeat expands (11, 13). We speculate that the polyglutamine tract is adjacent, in space, to the AXH dimerization domain. An expanded repeat may form aggregated local structures (e.g. β-sheet) that interfere with protein-protein and protein-RNA interactions, e.g. by blocking the respective binding sites, thereby affecting the normal cellular functions of ataxin-1.

REFERENCES

1. Skinner, P. J., Koshy, B. T., Cummings, C. J., Klement, I. A., Helin, K., Servedio, A., Zoghbi, H. Y., and Orr, H. T. (1997) Nature 389, 971–974.
2. La Spada, A. R., and Taylor, J. P. (2003) Neuron 38, 681–684.
3. Fernandez-Funes, P., Nina-Rosales, M. L., de Gouyon, B., She, W. C., Luchak, J. M., Martinez, P., Turiegos, E., Benito, J., Capovilla, M., Skinner, P. J., McCall, A., Canal, I., Orr, H. T., Zoghbi, H. Y., and Botas, J. (2000) Nature 408, 101–106.
4. Orr, H. T., and Zoghbi, H. Y. (2001) Hum. Mol. Genet. 10, 2307–2311.
5. Chen, H. K., Fernandez-Funes, P., Acevedo, S. F., Lam, Y. C., Kayment, M. D., Fernandez, M. H., Atkesh, A., Skoulakis, E. M., Orr, H. T., Botas, J., and Zoghbi, H. Y. (2003) Cell 113, 457–468.
6. Lin, X., Antalffy, B., Kang, D., Orr, H. T., and Zoghbi, H. Y. (2001) Nat. Neurosci. 3, 157–161.
7. Tevesian, S. G., Shih, H. H., Mendelson, K. G., Sheppard, K.-A., Paulson, K. E., and Yee, A. S. (1997) Genes Dev. 11, 383–396.
8. Sampson, E. M., Haque, Z. K., Ku, M.-C., Tevesian, S. G., Albanese, C., Pestell, R. G., Paulson, K. E., and Yee, A. S. (2001) EMBO J. 20, 4500–4511.
9. Lavender, P., Vandel, L., Bannister, A. J., and Kouzarides, T. (1997) Oncogene 14, 2721–2728.
10. Liu, K. M., Zhao, W.-G., Bhatnagar, J., Zhao, W.-D., Lu, J.-P., Simko, S., Schuenneman, A., and Austin, G. E. (2001) Lukemia (Baltimore) 15, 601–612.
11. Yue, S., Serra, H. G., Zoghbi, H. Y., and Orr, H. T. (2001) Hum. Mol. Genet. 10, 25–30.
12. Burritt, E. N., Davidson, J. D., Duvick, L. A., Koshy, B. Z., Zoghbi, H. Y., and Orr, H. T. (1997) Hum. Mol. Genet. 6, 513–518.
13. Hong, S., Kim, S.-J., Kim, S.-J., Park, Y., and Kang, S. (2002) Mol. Cell. Neurosci. 20, 298–306.
14. Kwon, K., Kim, S.-J., Park, Y., and Kang, S. (2003) Biochem. Biophys. Acta 1638, 35–42.
15. Miroux, B., and Walker, J. E. (1996) J. Mol. Biol. 260, 289–298.
16. van den Ent, F., Lockhart, A., Kendrick-Jones, J., and Low, J. (1999) Struct. Fold. Des. 7, 1181–1187.
17. Leslie, A. G. W. (1992) Recent changes to the MOSFLM Package for Processing Films and Image Plate Data. Int. CCP4/ESF-EACMB Newsletter Protein Crystallogr. 36, Daresbury Laboratory, Warrington, UK.
