Domain architecture of the polyglutamine protein ataxin-3: a globular domain followed by a flexible tail[1]

Laura Masino[a], Valeria Musi[a], Rajesh P. Menon[a], Paola Fusi[b], Geoff Kelly[a], Thomas A. Frenkiel[a], Yvon Trottier[c], Annalisa Pastore[a,*]

[a] *National Institute for Medical Research, The Ridgeway, London NW7 1AA, UK*
[b] *Università di Milano-Bicocca, Piazza della Scienza 2, 20126 Milano, Italy*
[c] *IGMBC, CNRS, INSERM, ULP, Université de Strasbourg, Illkirch Cedex 67404, France*

Received 5 May 2003; revised 12 June 2003; accepted 23 June 2003

First published online 23 July 2003

Edited by Irmgard Sinning

Abstract Anomalous expansion of a polyglutamine (polyQ) tract in the protein ataxin-3 causes spinocerebellar ataxia type 3, an autosomal dominant neurodegenerative disease. Very little is known about the structure and the function of ataxin-3, although this information would undoubtedly help to understand why the expanded protein forms insoluble nuclear aggregates and causes neuronal cell death. With the aim of establishing the domain architecture of ataxin-3 and the role of the polyQ tract within the protein context, we have studied the human and murine orthologues using a combination of techniques, which range from limited proteolysis to circular dichroism (CD) and nuclear magnetic resonance (NMR) spectroscopies. The two protein sequences share a highly conserved N-terminus and differ only in the length of the glutamine repeats and in the C-terminus. Our data conclusively indicate that ataxin-3 is composed by a structured N-terminal domain, followed by a flexible tail. Moreover, [^{15}N]glutamine selectively labelled samples allowed us to have a direct insight by NMR into the structure of the polyQ region.

Key words: Spinocerebellar ataxia type 3/Machado–Joseph disease; Huntington's disease; Triplet expansion; Polyglutamine disease; Spinocerebellar ataxia; Structure

1. Introduction

Spinocerebellar ataxia type 3 or Machado–Joseph disease (SCA3/MJD) is a member of the family of polyQ diseases, a group of neurodegenerative disorders caused by expansion of CAG trinucleotide repeats coding for polyQ in the gene products [1]. The polyQ region is the only feature shared by the polyQ disease proteins, that show otherwise no sequence homology and carry the polyQ stretch at different positions [2]. The disorders develop when the repeat length exceeds a threshold of 35–45 glutamines [1]. PolyQ expansions induce protein aggregation and the formation of neuronal intranuclear inclusions, a hallmark of polyQ diseases [1]. Increasing evidence suggests that protein misfolding and aggregation play a crucial role in pathogenesis [3,4]. A structural characterisation of the proteins involved in polyQ diseases would therefore provide insights into the mechanisms of neurodegeneration.

Human ataxin-3, the protein related to SCA3/MJD, is a ubiquitously expressed 41 kDa protein whose polyQ tract contains 12–40 glutamines in normal individuals and 55–84 glutamines in the pathogenic form [2]. Ataxin-3 was found in the genomes of several species, from nematodes to human, including plants [5]. Alignment of the family shows a conserved N-terminal block that corresponds to the sequence motif named Josephin (residues 1–198 in the human protein) [5]. The C-terminus, in contrast, is non-conserved throughout different species and contains long stretches of low complexity regions which include the polyQ tract, preceded by a highly charged region.

Structural modelling of ataxin-3 predicted a mainly α-helical structure, with the polyQ tract residing in a long helix [5]. The proposed structure is in partial agreement with the estimate of 35% α-helix, obtained by far-ultraviolet (UV) circular dichroism (CD) [6]. The predicted conformation for polyQ, however, is in contrast with most of the current hypotheses based on modelling works or on experimental evidence on ataxin-3 and model systems ([7] and references therein). These works suggest a random coil or a β-sheet conformation for solubilised polyQ tracts, and a β-sheet structure for polyQ aggregates. The determination of the three-dimensional (3D) structure of ataxin-3 could solve this debate.

In this article, we have structurally characterised native human ataxin-3, with the aims of defining its domain architecture and of investigating the conformation of the polyQ tract. Murine ataxin-3 was also studied. This orthologue has 88% sequence identity with the human protein, differing from it in the length of the polyQ tract (murine ataxin-3 has only six glutamines) and in the C-terminus of the molecule (Fig. 1A). Comparison of the two orthologues can therefore provide novel information on the behaviour of polyQ tracts of different lengths.

2. Materials and methods

2.1. Protein expression and purification

Murine ataxin-3 and two forms of non-expanded human ataxin-3, containing respectively 18-residue and 26-residue glutamine repeats, were produced. Human Q26 ataxin-3 was used for limited proteolysis experiments only, since the low yields obtained were insufficient for

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

[1] Supplementary data associated with this article can be found at doi:10.1016/S0014-5793(03)00748-8

0014-5793/03/$22.00 © 2003 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/S0014-5793(03)00748-8

structural studies. All proteins were expressed in *Escherichia coli* strain BL21 using the plasmids described below.

Human Q18 ataxin-3 cDNA was subcloned into a modified pET plasmid vector (Stratagene) to produce a glutathione S-transferase (GST)/ataxin-3 fusion protein, with a cleavage site for recombinant Tobacco Etch Virus (rTEV). The protein was purified using a glutathione Sepharose affinity matrix (Amersham Biosciences). Human Q26 ataxin-3 cDNA was obtained as described by Shehi et al. (submitted) and subcloned into a modified pET-11d plasmid vector. The hexahistidine-tagged protein was purified using Ni-NTA agarose (Qiagen) followed by anion exchange. Murine ataxin-3 cDNA was obtained as described in Shehi et al. (submitted) and subcloned into a pGEX6P-1 vector (Amersham Biosciences). The GST fusion protein was purified with glutathione Sepharose and cleaved from GST using PreScission protease. For purification purposes, two, 14 and five residues were added N-terminally to human Q18, human Q26 and murine ataxin-3, respectively.

Uniformly and selectively ^{15}N-labelled samples were produced by standard methods.

The final purity of all samples was higher than 95%, as assessed by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE) and mass spectrometry. Protein concentrations were determined using UV absorption, with calculated extinction coefficients at 280 nm of 27 670 M^{-1} cm^{-1} and 27 550 M^{-1} cm^{-1} for human and murine ataxin-3, respectively. Two features were consistently observed in all samples: the proteins were prone to degradation and had a tendency to aggregate non-specifically (although the degree of aggregation could be kept to a minor extent by careful sample handling) and to be adsorbed on quartz cuvette walls.

2.2. CD and fluorescence spectroscopies

CD measurements were performed on a Jasco J-715 spectropolarimeter at 20°C. Sample concentration was 2–4 μM and quartz cuvettes (Hellma) with path lengths of 1 mm were used. Spectral decomposition was achieved by combining the CONTIN, SELCON, and CDSSTR methods [8].

2-*p*-Toluidinylnaphthalene-6-sulphonate (TNS) (Sigma) fluorescence spectra were recorded at 20°C in quartz cuvettes using a SPEX FluoroMax fluorimeter. Excitation was at 320 nm and emission was scanned from 370 to 550 nm. The spectra were measured by adding human or murine ataxin-3 (to a final concentration of 2–4 μM) to a 30 μM TNS solution.

2.3. Nuclear magnetic resonance (NMR) spectroscopy

All NMR spectra were recorded at 600 and 800 MHz ^{1}H frequency at 25°C on Varian spectrometers. Experiments acquired included the one-dimensional (1D) spectrum, two-dimensional (2D) ^{15}N-HSQC spectra, a 2D ^{15}N-HSQC with refocussed INEPT periods and ^{1}H decoupling during t_1 to obtain natural linewidths in the ^{15}N dimension, a ^{15}N T_2-weighted HSQC with single fixed values of 10, 70 and 150 ms for the time allowed for T_2 relaxation, ^{15}N-(CLEANEX-PM)-FHSQC experiment [18] and a TROSY [19] at 800 MHz.

The NMR experiments were performed in different buffers, with ionic strength varying from 20 to 200 mM, and at pH 6 and 7.4. All the experiments led to similar conclusions.

2.4. Limited proteolysis

Thermolysin, trypsin, α-chymotrypsin, subtilisin and proteinase K were obtained from Sigma. Proteolysis was carried out at 25°C in 20 mM sodium phosphate, pH 8.0 (1 mM $CaCl_2$ was added to the thermolysin proteolytic mixture). Protein concentration was 0.9 mg/ml and the enzyme/substrate weight ratio was 1:100 for thermolysin, α-chymotrypsin, subtilisin and proteinase K, and 1:500 for trypsin. Proteolysis reactions were stopped by acidification with 0.1% trifluoroacetic acid (Fluka) and the proteolytic mixtures were separated by reverse phase high performance liquid chromatography (HPLC)

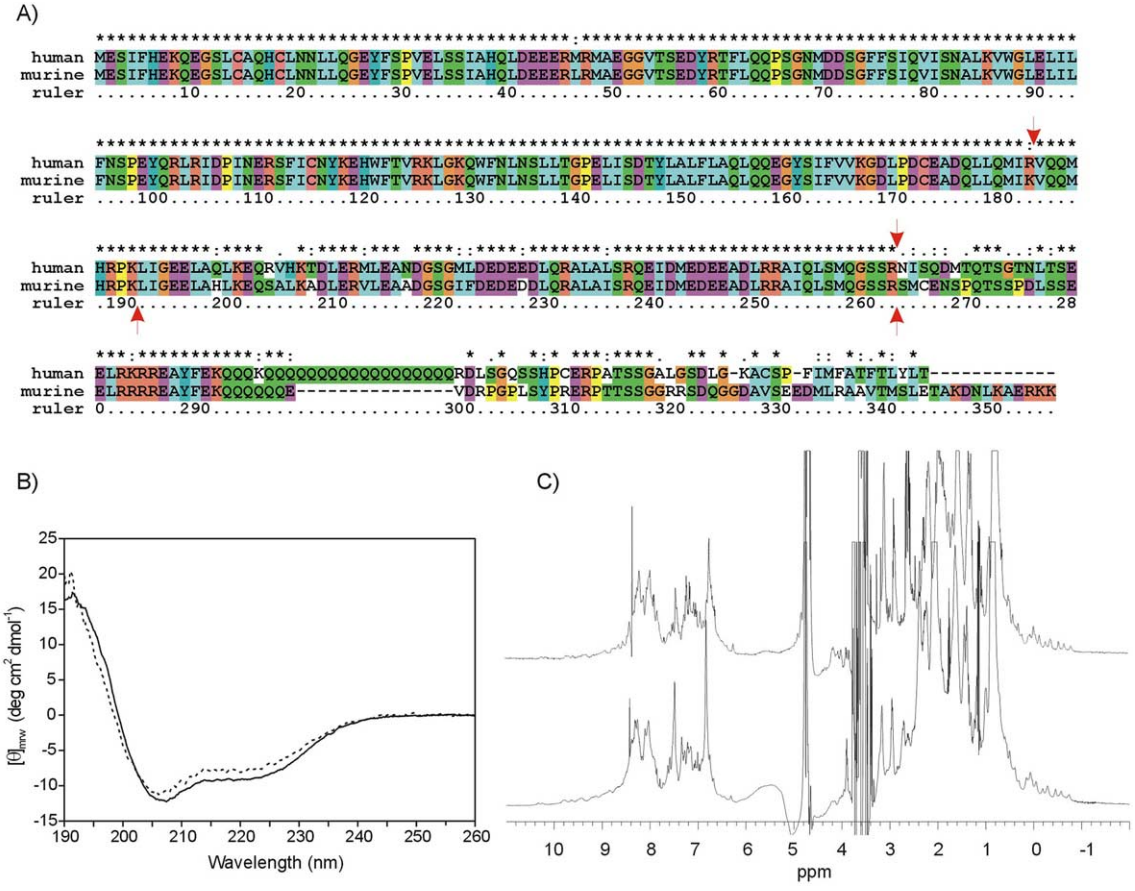

Fig. 1. A: Sequence alignment of human and murine ataxin-3. Sequence identity is indicated by stars. Decreasing degrees of sequence homology are indicated by semi-columns and dots. The numbering refers to the sequence of murine ataxin-3. Arrows indicate the two major proteolytic cleavage sites for each protein. B: Far-UV CD spectra of murine (dashed line) and human (continuous line) ataxin-3. C: 1D ^{1}H NMR spectra of murine (top) and human (bottom) ataxin-3.

using a C4 Vydac column (4.6 ×150 mm) (Separation Group, Hesperia, CA, USA). The fragments were analysed by SDS–PAGE, electrospray ionisation mass spectrometry and N-terminal sequencing by automated Edman degradation.

3. Results

3.1. Far-UV CD and NMR spectroscopies suggest that ataxin-3 is only partially folded

Human and murine ataxin-3 were analysed using CD and NMR spectroscopies. No significant differences were observed between the far-UV CD spectra of the human and murine orthologues (Fig. 1B). Spectral deconvolution [8] indicates a secondary structure content of 32% α-helix, 17% β-sheet, 20% β-turn, and 31% random coil, in agreement with the results obtained by Bevivino and Loll [6]. The significantly high percentage of random coil conformation estimated by this analysis suggests the presence of unstructured portions of the molecule alongside one or more folded regions.

The presence of a structured core in human and murine ataxin-3 was confirmed by 1D ^{1}H NMR spectra, as shown by the spectral dispersion over a range of 11 ppm and by the presence of resonances (very similar in the human and murine spectra) in the range from −1.0 to 0.8 ppm (Fig. 1C). Resonances in this region arise from protons in close contact with aromatic rings in structured conformations. Overall, the relatively sharp linewidth of the well dispersed peaks (~25 Hz) suggests the predominance of a monomeric species. Analytical ultracentrifugation and gel filtration data (not shown) are in agreement with this observation, although they do not provide conclusive results, probably because of the tendency of ataxin-3 to partial unspecific aggregation (see Section 2).

Two main resonance types are clearly distinguishable in the 2D ^{15}N spectra (HSQC) of murine and human ataxin-3 (Fig. 2A, B): well dispersed resonances typical of a folded conformation are present together with sharp peaks which overlap in the range 7.7–8.8 ppm (^{1}H dimension) and 115–126 ppm (^{15}N dimension). These chemical shift values are typical of a random coil conformation [9]. These results suggest that folded regions must coexist with unstructured tracts. The presence of regions with different mobility is further supported by T_2 relaxation experiments, which show selective disappearance of the well dispersed peaks first (see Additional materials). In addition to these two types, very broad resonances are also observable at chemical shifts typical of random coil species. Their intensities are not affected in TROSY experiments or by decreasing the pH, suggesting that they arise from regions of the proteins in a conformational exchange (see Additional materials).

The remarkable correspondence between the dispersed res-

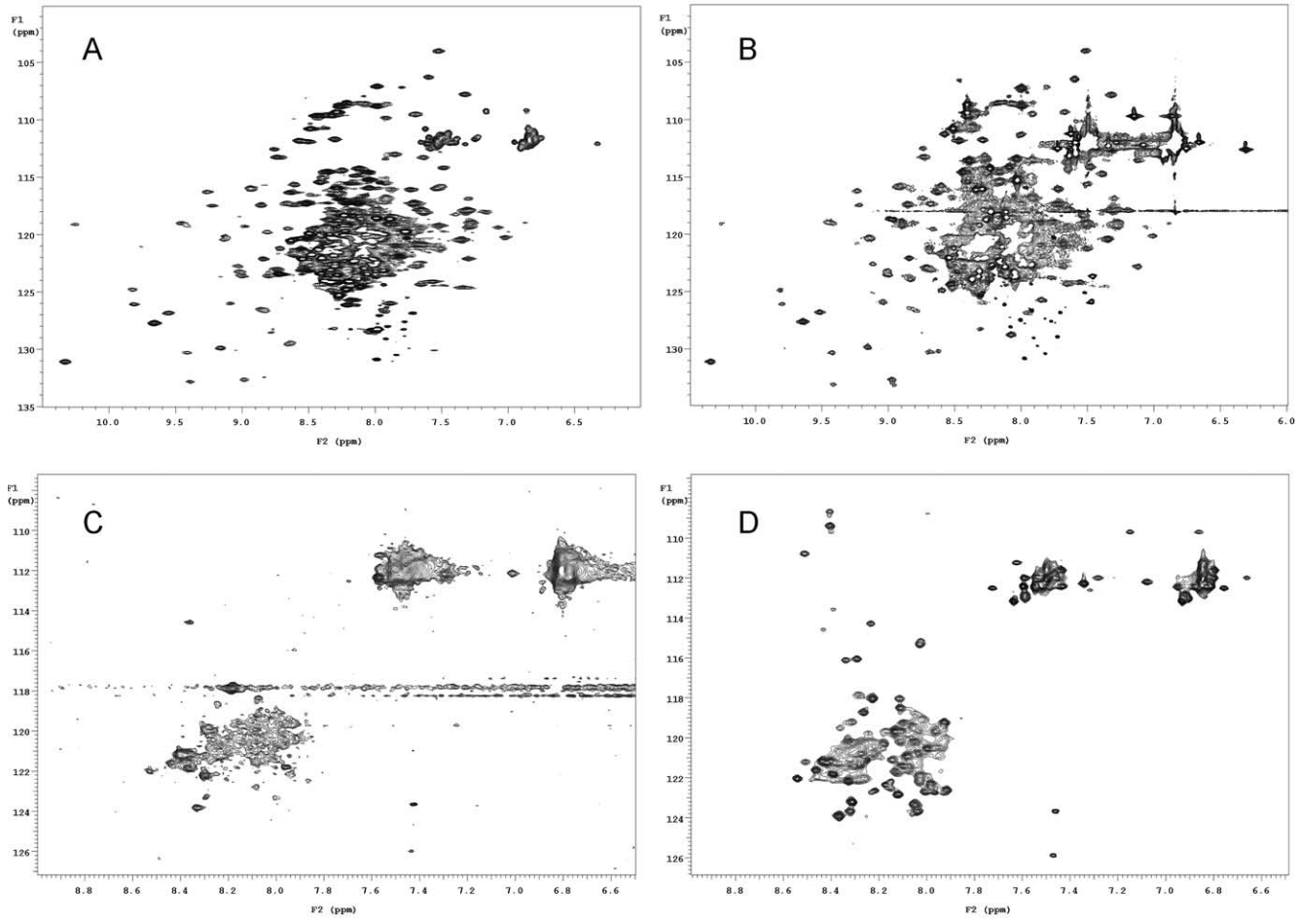

Fig. 2. HSQC spectra of uniformly ^{15}N-labelled murine (A) and human (B) ataxin-3. C: HSQC spectra of ^{15}N glutamine selectively labelled human ataxin-3 and, for comparison, D: enlargement of B plotted at a higher level.

onances in the human and murine spectra indicates that the folded regions must comprise the highly homologous portions of the two proteins and have the same conformation.

3.2. Limited proteolysis indicates the presence of a protease-resistant N-terminal domain

To probe the domain organisation and the level of compactness of ataxin-3, limited proteolysis experiments were performed on the murine and human proteins. Several proteolytic probes with different substrate specifities were used, namely trypsin, thermolysin, α-chymotrypsin, subtilisin and protease-K. The patterns of proteolytic digestion of human and murine ataxin-3 by trypsin, as detected by HPLC and SDS–PAGE, are shown in Fig. 3. For both proteins, the intact species are almost completely digested after 1 min of incubation and fragments of approximately 30 and 20 kDa are generated. As the proteolytic reaction proceeds, the 20 kDa fragments show remarkable resistance to degradation. Mass spectrometry analysis shows that the most stable species correspond to fragments −5–262 and −5–190 for murine ataxin-3, and −14–262, −14–190 and −14–182 for human ataxin-3 (negative numbers refer to the residues added N-terminally to human and murine ataxin-3 for purification purposes) (see Table in Additional materials). A similar pattern was obtained using the other proteolytic probes. Interestingly, the same pattern was also observed during spontaneous degradation of aged ataxin-3 samples (see Additional materials). These observations indicate that the N-terminal region of both proteins is more resistant to proteolytic cleavage, whilst the C-terminal region is significantly more prone to degradation. These conclusions fully agree with the NMR results and show that ataxin-3 comprises an N-terminal folded core and less compact regions at the C-terminus.

3.3. Glutamine selective labelling provides information about the polyQ structure in ataxin-3

The conformation of the C-terminal tail, which comprises the polyQ tract, was further investigated. The fluorescent probe TNS was used to detect the presence of exposed hydrophobic surfaces. A significant increase in TNS fluorescence and a blue shift of the emission maximum (440 nm) was observed when human ataxin-3 was added to TNS (see Additional materials). Murine ataxin-3 gave rise to a comparable effect, with a shift of the emission maximum to 449 nm. The data show that both proteins have hydrophobic residues that are exposed to solvent and may therefore be located in flexible regions of the proteins.

To identify the glutamine resonances, human ataxin-3 with ^{15}N selectively labelled glutamines was produced. Comparison of the HSQC spectra of uniformly and selectively labelled human ataxin-3 shows that most of the glutamine backbone amide resonances are concentrated within 8.5–7.9 ppm (^{1}H dimension) and 124–119 ppm (^{15}N dimension) (Fig. 2C, D). Although not all chemically equivalent, the resonances are grouped in a relatively narrow region. The chemical shifts, however, do not allow us to discriminate between α-helix, β-sheet or random coil conformation. The side chain amides are more informative, since they are all (with only one exception) concentrated between 7.5 ppm and 6.8 ppm (^{1}H dimension) and 112 ppm (^{15}N dimension). Comparison of the 1D ^{1}H spectra of murine and human ataxin-3 confirms that these peaks contain the contribution of the glutamine side chains in

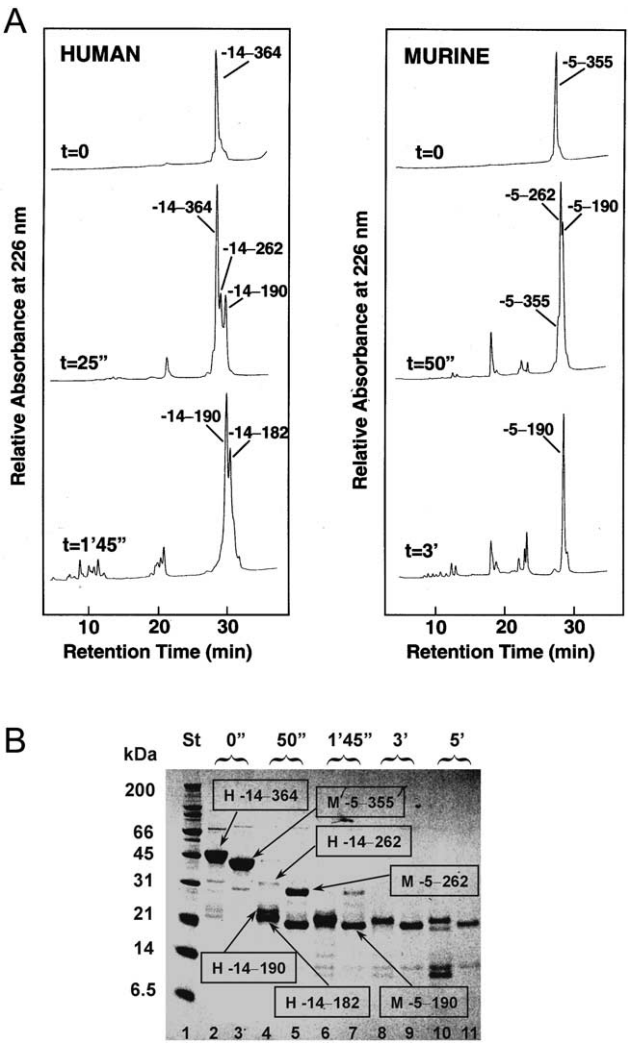

Fig. 3. A: HPLC analysis of proteolytic mixtures of human and murine ataxin-3 with trypsin at incubation time 0 (intact protein), 25″ and 1′45″, and 0, 50″ and 3′, respectively. B: SDS–PAGE analysis of the proteolytic mixture of human (lines 2, 4, 6, 8, 10) and murine (lines 3, 5, 7, 9, 11) ataxin-3 with trypsin. Aliquots from the reaction mixture were taken after 0, 50″, 1′45″, 3′ and 5′ of incubation. Molecular weight markers are shown in line 1.

the polyQ tract: the resonances at 7.49 and 6.83 ppm, that are barely visible in the spectra of the murine protein, are clearly observable as sharp peaks in human ataxin-3 (Fig. 1C). The equivalence of these resonances, with chemical shifts corresponding to random coil values, strongly suggests that the polyQ side chain amides experience a similar chemical environment. Further support to these observations was given by TOCSY and CLEANEX-PM experiments that provide information about flexibility and exposure to solvent (data not shown). The H_2N groups give rise to the only correlations observable in a TOCSY experiment performed with human ataxin-3 and have solvent exchange rates of the order of 12 s^{-1}. The data clearly show that the glutamines in the polyQ tract are flexible and exposed to solvent.

4. Discussion

Using different complementary techniques, we have characterised the domain architecture of ataxin-3. The possibility to

analyse the behaviour of the human and murine proteins in parallel proved essential in allowing us to reach our conclusions. Such a comparative genomic approach will undoubtedly prove of increasing importance in structural and functional genomic studies. Our results show that, in solution, ataxin-3 is composed of a structured and compact N-terminal domain followed by a more flexible C-terminus. These findings are in good agreement with the pattern of sequence conservation among different species, that indicates the presence of an N-terminal conserved motif, named Josephin (http://www.sanger.ac.uk/Software/Pfam/server). Indeed, the boundaries of the 20 kDa N-terminal fragment identified by limited proteolysis coincide with those indicated for this motif.

The protease accessible C-terminus comprises the polyQ tract. While it is difficult at this stage to identify conclusively in the NMR spectrum the glutamine backbone amides in the polyQ tract, comparison of ataxin-3 spectra with those of the GST-polyQ model systems [10] suggests interesting differences and similarities. In the GST-polyQ spectra, most of the observable resonances could be confidently attributed to the flexible polyQ tail, as those resonances arising from the GST moiety are severely broadened owing to efficient transverse relaxation in systems of this molecular weight ($\sim$56 kDa). In contrast, the spectrum of ataxin-3 appears crowded, indicating different hydrodynamic properties of ataxin-3 from GST-polyQ and that ataxin-3 is a non-compact monomeric species. Despite this difference, the chemical shifts of the glutamine side chains of ataxin-3 are remarkably similar both in their value and in the lack of dispersion to those observed for the GST-polyQ systems and are typical of a random coil conformation. This conclusion, together with the indication that the glutamine side chains are flexible, argues against the possibility of a β-conformation for the polyQ tract, stabilised by interactions across glutamine side chains [11]. Further insights into the conformation of polyQ will be provided by spectral assignment of the two orthologues. Production of the isolated N-terminal domain will also allow us to discriminate between glutamines in the polyQ tract and those in the rest of the molecule. Work along this line is currently in progress in our group.

While awaiting new data, we can draw some conclusions about the shape of the protein. Independent but complementary evidence suggests that, in solution, unbound ataxin-3 is a non-compact protein, as also shown by its tendency to undergo C-terminal proteolytic degradation and partial unspecific aggregation. This observation is in agreement with the elongated shape predicted for ataxin-3 by Bevivino et al. using analytical ultracentrifugation [6]. TNS binding and the chemical exchange observed in the NMR spectrum, two features typical of molten globule states, also suggest that, when not bound to other cellular components, the structure of ataxin-3 is much less compact than the one predicted by comparative modelling [5]. It is however possible that the protein may adopt a more compact structure only upon interaction with other cellular partners, which could stabilise the fold of the C-terminus. Ataxin-3 was shown to interact with the nuclear matrix [12,13] and to bind to histones and several transcriptional activators [14,15]. Moreover, ataxin-3 contains two ubiquitin-interacting motifs [16] and binds to the ubiquitin-like domain in the DNA repair proteins HHR23A and HHR23B [17]. Expansion of the polyQ tract could interfere with and change the interacting surfaces, thus preventing proper folding and promoting aggregation of the exposed glutamines. The structure determination of the folded domain of ataxin-3 and the further characterisation of its interactions with other cellular components will thus provide essential information on the protein's function and on the role of the polyQ expansion.

Acknowledgements: We wish to acknowledge D. Dalzoppo, S. Howell and E. Shehi for technical support, A. Fontana for constant encouragement, and P. Tortora for sharing results prior to publication.

References

[1] Zoghbi, H.Y. and Orr, H.T. (2000) Annu. Rev. Neurosci. 23, 217–247.
[2] Cummings, C.J. and Zoghbi, H.Y. (2000) Hum. Mol. Genet. 9, 909–916.
[3] Yang, W., Dunlap, J.R., Andrews, R.B. and Wetzel, R. (2002) Hum. Mol. Genet. 11, 2905–2917.
[4] Sanchez, I., Mahlke, C. and Yuan, J. (2003) Nature 421, 373–379.
[5] Albrecht, M., Hoffmann, D., Evert, B.O., Schmitt, I., Wullner, U. and Lengauer, T. (2003) Proteins 50, 355–370.
[6] Bevivino, A.E. and Loll, P.J. (2001) Proc. Natl. Acad. Sci. USA 98, 11955–11960.
[7] Masino, L. and Pastore, A. (2002) Biochem. Soc. Trans. 30, 548–551.
[8] Sreerama, N. and Woody, R.W. (2000) Anal. Biochem. 287, 252–260.
[9] Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) Biochemistry 31, 1647–1651.
[10] Masino, L., Kelly, G., Leonard, K., Trottier, Y. and Pastore, A. (2002) FEBS Lett. 513, 267–272.
[11] Perutz, M.F., Johnson, T., Suzuki, M. and Finch, J.T. (1994) Proc. Natl. Acad. Sci. USA 91, 5355–5358.
[12] Tait, D. et al. (1998) Hum. Mol. Genet. 7, 991–997.
[13] Perez, M.K., Paulson, H.L. and Pittman, R.N. (1999) Hum. Mol. Genet. 8, 2377–2385.
[14] Chai, Y., Wu, L., Griffin, J.D. and Paulson, H.L. (2001) J. Biol. Chem. 276, 44889–44897.
[15] Li, F., Macfarlan, T., Pittman, R.N. and Chakravarti, D. (2002) J. Biol. Chem. 277, 45004–45012.
[16] Hofmann, K. and Falquet, L. (2001) Trends Biochem. Sci. 26, 347–350.
[17] Wang, G., Sawai, N., Kotliarova, S., Kanazawa, I. and Nukina, N. (2000) Hum. Mol. Genet. 9, 1795–1803.
[18] Hwang, T.L., Mori, S., Shaka, A.J. and vanZijl, P.C.M. (1997) J. Am. Chem. Soc. 119, 6203–6204.
[19] Pervushin, K., Riek, R., Wider, G. and Wüthrich, K. (1997) Proc. Natl. Acad. Sci. USA 94, 12366–12371.