We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Prediction and Experimental Detection of Structural and Functional Motifs in Intrinsically Unfolded Proteins

Cesira de Chiara and Annalisa Pastore
MRC National Institute for Medical Research, The Ridgeway, London, UK

1. Introduction

Intrinsically unstructured proteins (IUPs) or proteins with intrinsically unstructured regions (IURs) have quickly gained increasing interest within the biological community because of their significant presence in the human genome and their potential links to major pathologies such as cancer, neurodegeneration and diabetes (Tompa, 2005; Tompa & Fuxreiter, 2008). The terms IUPs and IURs designate proteins or protein regions intrinsically devoid of a well defined tertiary structure. The concept was introduced a few years ago in the scientific literature as a brand new idea, which would represent a family of proteins thought to have been previously ignored or unappreciated (Dunker et al., 2001; Wright & Dyson, 1999). However, as in many other examples in Science, the concept of IUPs is far from being new. In the 70s, it was universally accepted that what were known as 'biologically active peptides' had no intrinsic structure, often being too short to have a proper hydrophobic core. Peptides would/could however fold in a definite conformation upon interaction with a partner/receptor, thus having all the features of modern IUPs (Boesch et al., 1978). Hormones and opioid peptides are two among several of the best studied examples.

The concept of intrinsic disorder and/or flexibility has now been extended to proteins and has deeply transformed our perception of the importance of protein dynamics as opposed to the static picture introduced by years of crystallographic studies. Even more important is the fact that accepting the existence of IUPs proposes a unique paradigm in which function can be directly linked to structural disorder rather than to a defined structure.

IUPs have been classified in two broad categories. In the first family, IUP's function is achieved through binding to one or several partner molecule(s) in a structurally adaptive process, which enables an exceptional plasticity in cellular responses. These proteins do not form a structure by themselves and are functionally inactive in the absence of a partner, but structure can be induced upon recognition of another molecule. When bound to a substrate, they are able to acquire a structure and become rigid, according to an induced-fit mechanism or to what has been recently generalized in the concept of 'conformational fuzziness' (Wright & Dyson, 2009). Macromolecular association rates have in fact been demonstrated to be highly enhanced by a relatively non-specific association enabled by flexible recognition segments. Molecular recognition occurring in this way has been

described according to the 'fly-casting' (Shoemaker et al., 2000) or 'protein fishing' (Evans & Owen, 2002) mechanisms. Examples of proteins belonging to this family are those bearing RNA binding motifs which acquire a structure only upon interaction to RNA.

In the second family, IUPs work as entropic chains exploiting their ability of fluctuating over ensembles of structural states with similar conformational energies. In this way they either generate force against structural changes or influence the orientation/localization of attached domains (Dunker et al., 2002). According to these properties, they are active in their unstructured form and play the role of flexible linkers necessary to allow other portions of the protein to move like 'a dog on a leash' and ultimately to interact with other partners. A classical example of such a case is the IUR, called PEVK, of the muscle protein titin (Labeit & Kolmerer, 1995). This region confers some of the passive elastic properties of the titin filament, providing the stiffness required in muscle contraction to keep the sarcomere in register (Greaser et al., 2000).

These unique features are exploited in many biological processes thus explaining the multiplicity of different functions in which IUPs are involved (Dunker et al., 2002). Protein disorder prevails for instance in signaling, regulatory and cancer-associated proteins. The functional importance of protein disorder is also underlined by its dominant presence in proteins associated with signal transduction, cell-cycle regulation, gene expression and chaperone activities (Dunker et al., 2002; Iakoucheva et al., 2002; Tompa & Csermely, 2004; Uversky, 2002; Ward et al., 2004b). Because of their susceptibility to degradation, IUPs have also been linked to the ubiquitin (Ub)/proteasome pathway (Csizmok et al., 2007).

In this chapter, we discuss the problems related to the prediction, production and characterization of IUPs/IURs. We take as a representative case study ataxin-1, a human protein of biological and clinical interest that is related to the neurodegenerative disease spinocerebellar ataxia of type-1. Using ataxin-1, we retrace how the application of different bioinformatics tools has contributed to shed light on the structure and functions of the protein since its first identification. We also provide an update on the physico-chemical methods used to translate the sequence information into structural and functional models of the protein and its interactome. This hands-on example might provide a valuable paradigm of how correct identification of linear motifs and IURs can provide the key for understanding protein function.

2. Prediction methods of IUPs

While it has become increasingly easy to appreciate the importance of IUPs/IURs, their prediction and experimental characterization remain somewhat problematic. Since Romero et al. (Romero et al., 1997) indicated for the first time that lack of a defined protein tertiary structure is predictable on the sole basis of the primary sequence, several different methods have been developed that enable prediction (reviewed in Radivojac P et al. (Radivojac et al., 2007) (**Table 1**). They are based on different definitions of IURs and detect different indicators such as hydrophobicity, sequence composition, secondary structure content, etc.

These programs, however, are not always entirely reliable. Most weaknesses arise from the intrinsic difference between the conceptual and operational definition of IUP/IUR. As mentioned above, conceptually, there are two classes of IUP/IUR. The first does not form a structure by itself, but this can be induced by a partner. Proteins belonging to the second class can perform their function in three ways: (1) through the newly acquired structure, (2) by inducing structural changes in their partners and modulating the function of these

partners, or (3) through formation of protein complexes with partners. This class of IUP/IUR does not function in its unstructured form, i.e., the unstructured form is inactive. The second class performs its cellular function without forming a structure: the unstructured form is the functional one.

Software	Prediction	Based on
Disopred2 (http://bioinf.cs.ucl.ac.uk/disopred) (Ward et al., 2004b)	Regions lacking ordered secondary structure	Cascaded support vector machine classifiers trained on PSI-BLAST profiles
PONDR (http://www.pondr.com/) (Radivojac et al., 2003)	Non rigid regions from random coils to partially unstructured regions and molten globules	Local aminoacid composition, flexibility, hydropathy, etc
GlobPlot (http://globplot.embl.de/) (Linding et al., 2003b)	Regions with high propensity for globularity based on propensity for secondary structure and random coil	Russell/Linding scale of disorder
DisEMBL (http://dis.embl.de/) (Linding et al., 2003a)	LOOPS (region without secondary structure) HOT LOOPS (loops with high mobility) REMARK465 (regions of crystal structures lacking of electron density)	Neural networks trained on crystal structures
IUPred (http://iupred.enzim.hu/) (Dosztanyi et al., 2005a, 2005b)	Regions devoid of well-defined 3D structure under native conditions	Local aminoacid composition used to estimate inter-residue interaction energy

Table 1. List of predictors of protein disorder used for ataxin-1 in this study. The table, which illustrates the features of the different methods, is adapted from Ferron et al. (Ferron et al., 2006).

The operational definition used by bioinformatic software is typically based on the likelihood of the protein/peptide forming a structure under certain (often poorly defined) cellular environments, with little or no information on how the protein/peptide functions or what partner the IUP/IPR may have. Intrinsically unstructured proteins are for instance characterized by a low content of bulky hydrophobic amino acids and a high percentage of polar and charged amino acids. As a consequence, they do not contain enough residues to build a hydrophobic core that is typical of stably folded globular proteins. Another symptomatic indication of an IUP is the presence of low complexity motifs, i.e. sequences with over-representation of just a few residues (e.g. polyglutamine stretches, arginine-glycine (RG) repeat observed in RNA binding proteins or arginine-serine (RS) repeats observed in splicing factors). Of course, while low complexity sequences are a strong indication of disorder, the reverse is not necessarily true, that is, not all disordered proteins have low complexity sequences.

We can then expect that as these predictions become more robust the more we shall be able to distinguish between the different IUPs/IURs and have functional clues. For the time being, this software should be handled with care. For instance it is quite important to use more than one approach on the same sequence and compare the resulting scores.

3. Detection of linear functional motifs

It is difficult to underestimate the importance of the concept of structural motifs or modules in the world of globular proteins (Konagurthu & Lesk, 2010; Schaeffer & Daggett, 2010). A similar role is taken now by linear motifs in the world of IUPs. Eukaryotic linear motifs (ELMs) are short stretches of eukaryotic protein sequences (typically 3-10 amino acids long) to which a molecular function is associated. These segments provide regulatory functions independently of protein tertiary structure. While also potentially present in stably folded proteins, they acquire a particular importance in IUPs because in these they cannot be shielded by structure. It is in fact found experimentally that short functional sites, which are frequently involved in regulatory processes, must reside in non-structured or non-globular regions or, when within globular domains, in flexible highly exposed loops (Gibson, 2009). Examples of linear motifs are phosphorylation sites, nuclear localization signals or signalling sequences. A useful tool for the prediction of linear motifs is the ELM database (http://elm.eu.org/) developed from a collaborative effort between EMBL and University of Rome (Diella et al., 2008; Puntervoll et al., 2003). A sequence of interest can be screened against this database to quickly suggest the position of ELMs.

4. When can we be sure that a protein is a IUP?

Many of us have been confronted with natural or recombinant proteins that, after purification, result in being unfolded even when they are expressed in a soluble form. Are these bona fide IUPs? It is some times difficult to distinguish between IUPs and orphan complex proteins, yet there is a profound difference between the two families. IUPs have no intrinsic capability to adopt a definite structure at least in the absence of a partner. Their conformational state is also independent of the way they have been purified. On the contrary, Orphan complex proteins are able to fold but their tertiary fold is only marginally stable even though not being intrinsically devoid of a 3D structure (Sjekloća et al. 2011). This is often due to the absence of a partner which stabilizes their fold. These proteins might be produced as unfolded proteins, as folded/unfolded mixed populations or as species highly prone to aggregation. Their misbehaviour can in principle be reduced or neutralized by finding more suitable conditions or less drastic purification protocols (e.g. concentration is often a problematic step). The only guidance to distinguish among the two families may be sequence analysis coupled to comparison with other homologous. If the motif is observed in other proteins where it adopts a stable fold, it is difficult to believe that it is intrinsically unfolded. In such a case, we suggest putting more care in protein production and attempting to identify the partners able to provide stabilization. A way to solve the ambiguity when the protein of interest is purified as an unfolded monomer can be to register an NMR spectrum directly on the cell lysate, if proper overexpression can be achieved. This circumvents the purification step and 'shows' the structural properties of the protein independently of human intervention.

5. Experimental methods to study IUPs

How can we study the structure of intrinsically unstructured molecules? The statement sounds like an oxymoron in which the two terms contradict each other. The problem closely resembles the difficulties encountered in the study of denatured states of ordered proteins. It goes without saying that IUPs cannot be crystallized in isolation because of their flexibility. Solution studies are therefore more appropriate to characterize their structural states. The ultimate goal of these studies cannot, however, be that of obtaining *the* structure but rather to describe the protein as an ensemble of rapidly interconverting alternative structures characterized by differing backbone torsion angles.

Among the biophysical methods able to provide structural and dynamic information, two techniques are probably the ones that have given more interesting results over the last few years: Nuclear magnetic resonance (NMR) and Small Angle X-ray Scattering (SAXS). Despite the disadvantages of the very poor chemical shift dispersion (particularly for proton and aliphatic carbon resonances) because of conformational averaging, NMR is particularly powerful thanks to its ability of measuring different independent observables. These include secondary chemical shifts, residual dipolar couplings (RDC), hydrogen bonds, torsional angles and long-range NOE upon spin labelling (for an exhaustive review see Mittag & Forman-Kay, 2007). Among them, detection of residual secondary chemical shifts is probably the simplest qualitative way to detect complete or partial local disorder of a chain. Most of the other NMR observables can be exploited in a more quantitative way: since the early implementation of the ensemble-averaged nOe distance restraints (Bonvin & Brunger, 1996), a variety of restraining algorithms, including simultaneous time and ensemble averaging (Fennen et al., 1995), have been developed and have been used to describe native, transition, intermediate, and unfolded states (Clore & Schwieters, 2004a, 2004b; Kuszewski et al., 1996; Vendruscolo & Paci, 2003; Vendruscolo & Dobson, 2005). A modern approach that takes into account protein flexibility is that of imposing penalizing forces if the calculated average distances at a given time across an ensemble of simulated molecules (the 'replica ensemble') do not match the experimental ones. It is interesting to note that RDCs provide a particularly powerful way to assess protein structures using an absolute reference system despite the original scepticism towards applying these measurements for the treatment of IUPs and IURs which pose fundamental problems in the way structural averaging should be used. To address the problem, new methods based on RDC measurements have been developed to provide detailed information on protein dynamics also in cases of conformational averaging, for instance based on analytical deconvolution, Gaussian axial fluctuation methods and restrained molecular dynamics simulations. Most of these methods have, however, been used to assess the dynamics of relatively small amplitude methods. Their application to flexible and conformationally interconverting molecules such as IUPs and IURs remains to be fully established.

SAXS, by which it is possible to achieve the measurement of molecular dimensions and a description of the overall shape of the ensemble, is, in many ways, complementary to NMR. A key advancement to quantitatively characterizing flexible proteins in solution by SAXS was achieved with the implementation of the approach known as ensemble optimization method (EOM) (Bernado et al., 2007). In this approach, flexibility is taken into account by postulating the coexistence of different conformational states for the protein contributing to the experimental scattering pattern. The different conformers can then be selected by genetic algorithms from a pool containing a large number of randomly generated models covering the protein configurational space. The EOM selected models are then analysed by quantitative statistical criteria also developed to determine the optimal number of

conformers necessary to represent the ensemble. When possible, the quality of the analysis is increased by simultaneous fitting of multiple scattering patterns from deletion mutants, a procedure which is somewhat equivalent to improving the quality of sequence alignment introducing information about sequence homologues. The EOM protocol has now been validated by applying it to the study of several examples of completely or partially unfolded proteins and on multidomain proteins interconnected by flexible linkers and has shown to be a robust and helpful approach to the study of IUPs.

6. Ataxin-1 as a case study

To discuss the problems related to the prediction and characterization of IURs we chose as a representative case study ataxin-1, a human protein of biological and clinical interest that is related to the neurodegenerative disease spinocerebellar ataxia of type-1 (SCA1). This hereditary pathology is dominant and a member of a small family of neurodegenerative diseases linked to protein aggregation and misfolding, all caused by anomalous expansion of polyglutamine (polyQ) tracts in the gene coding region (Orr & Zoghbi, 2007) (**Figure 1**).
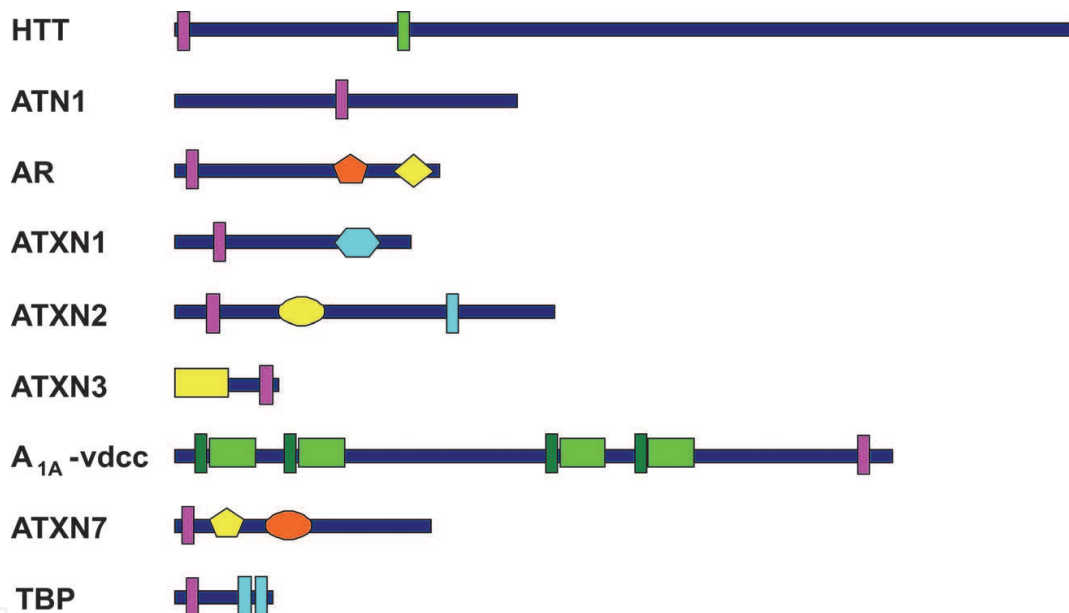

Fig. 1. Schematic representation of the protein members of the polyQ disease family. The sequence predicted to be intrinsically unfolded according to predictor DisEMBL (Radivojac et al., 2003) is indicated by a blue line. The position of polyQ tracts is indicated with magenta boxes. The position of other structural and functional domains, as reported by SMART (Schultz et al., 1998) and Pfam (Finn et al., 2010) databases, is indicated with geometric symbols of different colours. HTT: Huntingtin; ATN1: Atrophin 1; AR: Androgen receptor; ATXN1: Ataxin-1; ATXN2: Ataxin-2; ATXN3: Ataxin-3; A_{1A}-vdcc: A_{1A}-voltage dependent calcium channel; ATXN7: Ataxin-7; TBP: TATA-box binding protein.

Other members of the polyQ pathology family are Huntington's chorea, Machado-Joseph disease and other spinocerebellar ataxias. Elongation of polyQ tracts is the result of a gene polymorphism where unstable consecutive CAG triplets may become expanded during DNA replication. The mutated proteins form intracellular aggregates which are thought to be toxic for the cell and lead to cell death according to a gain rather than loss of function mechanism. While generally accepted that polyQ expansion is the leading factor in

triggering pathology, it has become increasingly clear that protein context is an important element that modulates protein behaviour and pathology. Despite best efforts, all the polyQ pathologies are currently incurable. A way to develop a specific treatment is through the identification of the function(s) of the native proteins. In this endeavour, the support of bioinformatics analysis of the protein sequence becomes essential to predict structure and a great help in suggesting a function for these proteins. The sequences of the polyQ disease families are very different in length and position of the polyQ tract. They are also not homologous and share only a rather loose common feature: they all seem to contain large IURs which are sometimes interrupted by readily identifiable globular domains. As such, they constitute an excellent example to discuss the problems inherent in detecting and studying IUPs/IURs.

In the following sections we shall discuss a detailed analysis of the ataxin-1 primary sequence and how this study has suggested working hypotheses which could then be tested experimentally using structural and cellular approaches. We discuss the identification of two major potential IURs in ataxin-1 which contain short linear motifs important for phosphorylation, aggregation and protein-protein interactions that are directly related to pathogenesis (Chen et al., 2003; de Chiara et al., 2009; Emamian et al., 2003; Jorgensen et al., 2009; Klement et al., 1998). We pay particular attention to the prediction and characterization of the structural and functional features of the different protein regions and to the identification of key ELMs of crucial importance both for the normal and anomalous behaviour of the protein.

6.1 Ataxin-1 domain architecture

Ataxin-1 is a ca. 98 kDa protein which is well conserved through vertebrates, ubiquitously expressed and mainly localised in the cell nucleus (Klement et al., 1998). As with all the other members of the polyQ protein family, which it shares a common pathogenic mechanism with, ataxin-1 was originally identified by gene mapping (Banfi et al., 1994). In its non-expanded form, human ataxin-1 is a protein of 816 amino acids although expansion can significantly increase its length (**Figure 2**). Different predictions of IURs suggest the presence of long unstructured regions (**Figure 3**).
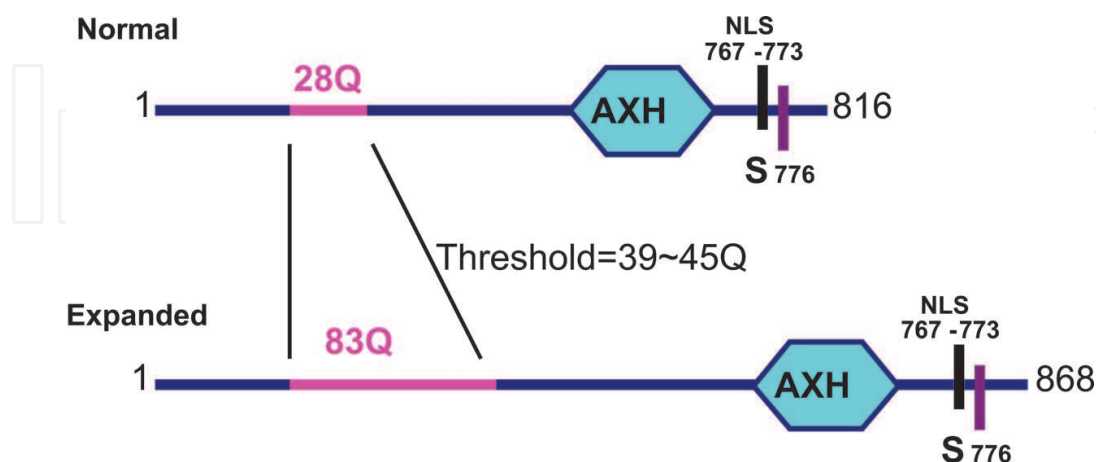

Fig. 2. Architecture of non-expanded and expanded ataxin-1. The polyQ tract is shown in magenta and the AXH domain in cyan. The positions of the nuclear localization signal (NLS) and a phosphorylation site important for protein interactions and for pathology are also indicated (de Chiara et al., 2009).
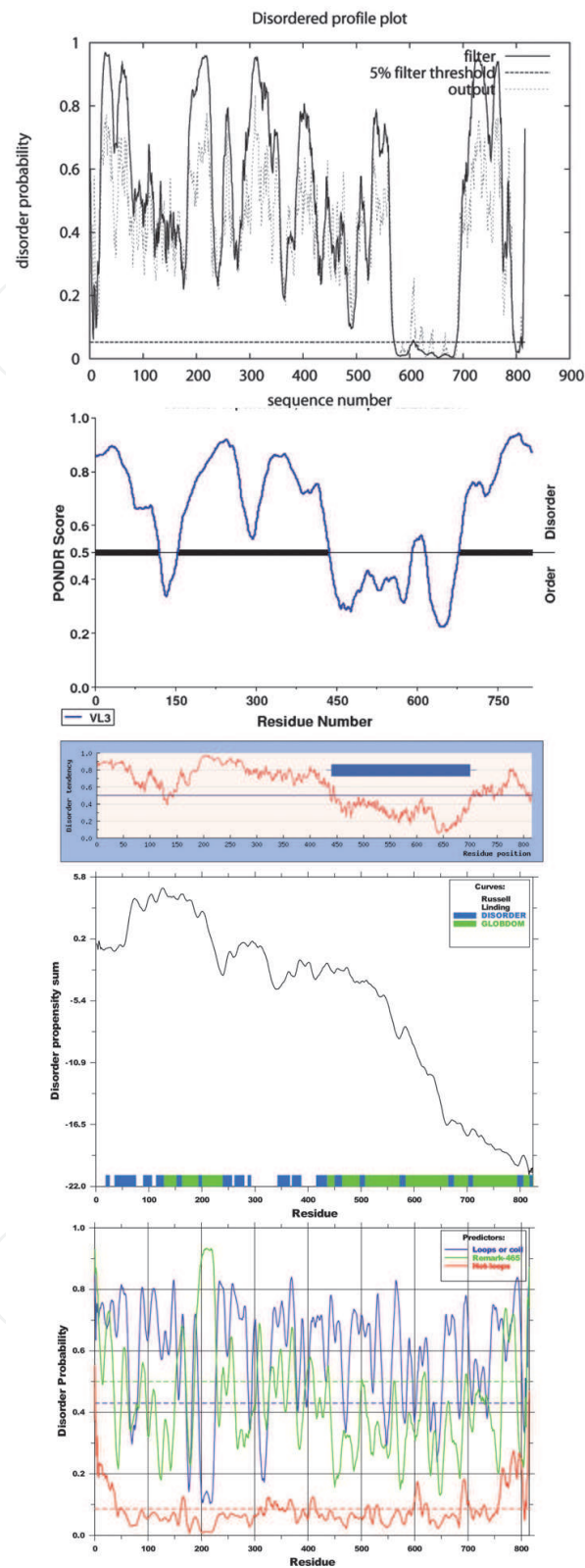
Fig. 3. IUR regions of ataxin-1 according to different predictors. From top to bottom: Disopred2 (Ward et al., 2004a; Ward et al., 2004b), PONDR (Radivojac et al., 2003), IUPRED (Dosztanyi et al., 2005a, 2005b), GlobPlot (Linding et al., 2003b) and DisEMBL (Linding et al., 2003a).

6.1.2 The N-terminal IUR

6.1.2.1 The N-terminal IUR according to different IUR predictors

In addition to the low complexity polyQ tract, there is a long region potentially unfolded from the N-terminus up to the AXH domain (567-689) which seems to be the only region of the protein that is autonomously folded (**Figure 2**). According to GlobPlot (Linding et al., 2003b), for instance, the region of ataxin-1 from the N-terminus to the beginning of the AXH domain is predicted as a series of multiple disorder regions. Among these are short low-complexity sequences (47-64, 88-99, 154-169, 366-377) as defined by SMART (Letunic et al., 2009; Schultz et al., 1998). Disopred2 (Ward et al., 2004a; Ward et al., 2004b), PONDR (Radivojac et al., 2003), DisEMBL (Linding et al., 2003a) and IUPRED (Dosztanyi et al., 2005a, 2005b) also concordantly predict this region as almost entirely disordered (**Figure 3**).

6.1.2.2 The polyQ tract

The N-terminus of the human protein is characterized by the presence of a highly polymorphic almost uninterrupted polyQ stretch which ranges from 4 to ~39 Qs in normal population and is expanded to ~40-83 Qs in SCA1-affected individuals (Zoghbi & Orr, 2008). Pathology typically develops when the repeat length exceeds a threshold of 35-45 glutamines (Genis et al., 1995; Jayaraman et al., 2009; Orr et al., 1993). Expansion of this region is a feature not shared among other species suggesting an evolutionary gain associated only to humans. The polyQ tract of ataxin-1 starts at residue 197. Indeed, the structure of polyQ stretches in solution has been shown experimentally both by CD and NMR spectroscopy to be a random coil when in a non-aggregated form (Masino et al., 2002). This is at variance with predictions by SMART which propose a helical coiled-coil region for the same region (amino acids 193-230). The discrepancy should anyway be ascribed by a bias in SMART for poly-amino acids. Interestingly, we now know that expansion of the polyQ tract in ataxin-1 is a condition necessary but not sufficient for triggering disease: two other motifs, a nuclear localization signal (NLS) and phosphorylation of S776, both located at the C-terminus and discussed in a session below, have been proved to also be required. As for the other polyQ proteins the native function of the polyQ tracts is unknown, although their presence has mostly been detected in proteins associated with transcriptional regulation activity. Indeed, the transcriptional regulator poly-Q binding protein-1 (PQBP1) has been found to bind ataxin-1 in a polyQ length-dependent manner, suggesting that PQBP1 and mutant ataxin-1 may act cooperatively to repress transcription and induce cell death (Okazawa et al., 2002). Direct evidence to support this hypothesis is now necessary.

6.1.2.3 Prediction of significant ELMs in the N-terminal IUR

Although ataxin-1 has been reported as a protein shuttling in and out the nucleus and the cytoplasm, there is a large body of evidence showing that the protein is predominantly located in the nucleus. Restricting the search for candidate short linear motifs and post-translational modifications within the N-terminal IUR to those significant for the nuclear localization, several ELMs have been predicted by using the ELM resource (Diella et al., 2008; Puntervoll et al., 2003) (**Table 2**).

A plethora of potential phosphorylation sites are predicted by the ELM resource in the N-terminal IUR of ataxin-1, among which only Ser239 (Vierra-Green et al., 2005) and Ser254 have been experimentally verified (Dephoure et al., 2008). Phosphorylation of these two serine residues supports the prediction by ELM of two candidate Class IV WW domain interaction motifs present in the regions 236-241 and 251-256 which mediate phosphorylation-dependent interactions. In addition to the WW domain motifs, other

ELM	Sequence	Position	Description	Experimentally Verified
LIG_14-3-3_1	RRWSAP	773-778	Mode 1 interacting phospho-motif for 14-3-3 proteins	Yes
LIG_14-3-3_3	KAPTLP RTASPP KAESSR HSASEP	29-34 260-265 309-314 439-444	Interacting phospho-motif for 14-3-3 proteins not matching Mode 1 and Mode 2	
LIG_FHA_1	EGTAWLP PVTSAVA PHTLTLG PGTQPLL FVTTALP	41-47 154-160 279-285 486-492 523-529	Phosphothreonine motif binding FHA with preference for large aliphatic at pT+3	
LIG_FHA_2	AGTSVEL LSTGLDY QATHREA PSTLNDK GSTDMEA FLTKIEP	63-69 81-87 400-406 408-414 496-502 758-764	Phosphothreonine motif binding FHA with preference for acidic aminoacid at pT+3	
LIG_WW_4	TPGSPP ISSSPQ	236-241 251-256	Phosphopeptide motif interacting with Class IV WW domain	
MOD_PKA_1	RRWSAPE	773-779	PKA-type AGC kinase phosphorylation site	Yes[a]
MOD_ProDKin_1	TPGSPPP ISSSPQN	236-242 251-257	MAPK phosphorylation site in higher eucaryotes	Yes[b] Yes[b]
MOD_CK2	RRWSAPE	773-779	CK2 phosphorylation site	
MOD_GSK3_1	ISSSPQNT RRWSAPES	251-258 773-780	GSK3 phosphorylation recognition site	Yes[b]
LIG_ULM_U2AF65_1	RKRRWS	771-776	Pattern in ULMs of SF1 and SAP155 which bind to the UHM of U2AF65	Yes
LIG_USP7_1	PATSR AGTSV PVTSA AVASA AESSR PYESR PSPSD ASPST PVGST PKPSL	20-24 63-67 154-158 158-162 310-314 356-360 366-370 406-410 494-498 795-799	Variant of the USP7 NTD domain based on the MDM2 and P53 interactions	
MOD_SUMO	PKSE LKTE	529-532 594-597	Motif recognized for modification by SUMO	
TRG_NLS_MonoCore_2	TRKRRW	770-775	Monopartite variant of the classical basically charged NLS. Core version	Yes

Table 2. Prediction of relevant functional linear motifs in the N- and C-terminus of human ataxin-1 by the ELM database (http://elm.eu.org/) (Diella et al., 2008; Puntervoll et al., 2003). Among all the predicted phosphorylation sites only the ones which have been

confirmed in vivo and are supportive of the prediction of other phosphopeptide motifs have been included. [a]Phosphorylation site and kinase experimentally confirmed in cerebellum (Jorgensen et al., 2009). [b]Phosphorylation site confirmed, kinase not identified (Dephoure et al., 2008; Vierra-Green et al., 2005)

putative phosphorylation-dependent protein-protein interaction motifs are predicted in the N-terminus of ataxin-1. Among these are several forkhead-associated FHA domain type-1 and -2, and adaptor protein 14-3-3 ligand motifs suggesting that the N-terminus of ataxin-1 may play a significant role in the assembly of the protein interactome.

6.1.2.4 Self association region (SAR)

At the cross-point between the N-terminal IUR and the AXH domain, a region of ca 100 aa (495-605) was identified in a yeast two hybrid system as responsible for protein self association (SAR) in cell (Burright et al., 1997). SAR shares 39 aa with the N-terminus of the AXH domain (567-689) which has been shown to be dimeric in solution (Chen et al., 2004; de Chiara et al., 2003). Therefore, this region seems to account for dimerization of the full-length ataxin-1. Interestingly, according to PONDR (VL3 predictor) and IUPRED, both based on the analysis of the local aminoacid composition, the full region ~440-700, which includes SAR and the AXH domain, is predicted as a potentially folded domain (**Figure 3**).

6.1.3 AXH domain

6.1.3.1 The structure of the ataxin-1 AXH domain

Soon after gene identification (Banfi et al., 1994), the analysis of the ataxin-1 sequence and the prediction of the secondary structure performed from multiple alignment of the protein from different species allowed the discovery of a new small putative independently folded domain (ca. ~130 aa) (with predicted predominantly beta structure) (de Chiara et al., 2003). The domain, successively named AXH (for Ataxin-1 Homology domain), did not show any detectable homology with any other known folding units (SMART accession number SM00536; http://smart.embl-heidelberg.de/) (Letunic et al., 2009; Schultz et al., 1998). A few years later, the homology between the ataxin-1 AXH and a region of an unrelated protein, the transcription factor HBP1, was detected (Mushegian et al., 1997). The two proteins share ~28% identity and ~54% similarity with the HBP1 AXH domain showing a ca. 10 aminoacids insertion loop between secondary structures respect to the ataxin-1 domain (de Chiara et al., 2003) (**Figure 4**).

The structure of ataxin-1 AXH, as solved by X-ray crystallography, consists of a non-canonical oligonucleotide- and oligosaccharide-binding (OB) fold (Chen et al., 2004; Murzin, 1993) (**Figure 5**). The AXH appears as a constitutive asymmetric dimer which crystallizes as an asymmetric dimer of dimers. Each monomer displays a common structure in the C-terminal part (residues 610-685), recognizable as the OB-fold, which superposes with an average root mean square deviation of 0.90 ± 0.06 Å on the backbone atoms. Conversely, approximately the first 30 N-terminal aminoacids show appreciable main chain differences between each of the two monomers in the dimer, with the same stretch of aminoacids adopting alternative secondary structures in the two cases. In this respect, the AXH domain represents an interesting example of a chameleon protein, which is a protein that adopts different folds under different environments. Interestingly, the observed structural differences in ataxin-1 are not induced by different experimental conditions or by the presence of ligands. Instead, they are present in the context of the same protein.
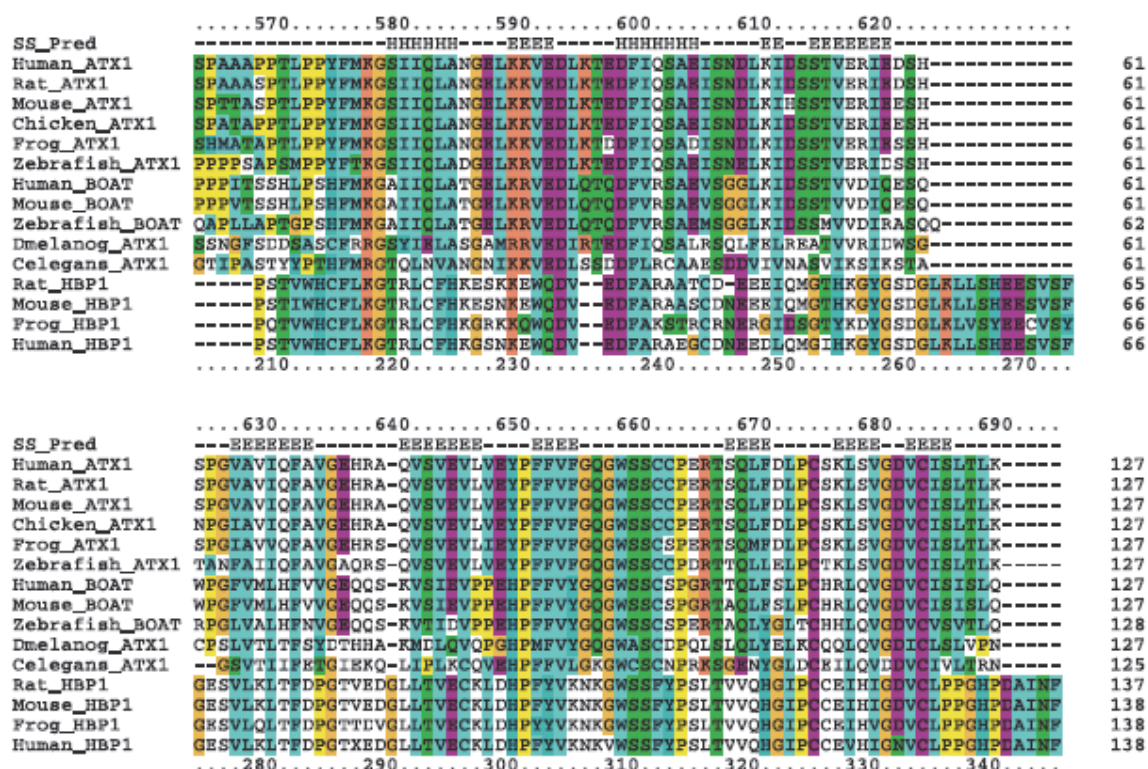
```
         .....570.......580.......590.......600.......610.......620........
SS_Pred  --------------------HHHHHH----EEEE-----HHHHHHH------EE--EEEEEEE--------------
Human_ATX1     SPAAAPPTLPPYFMKGSIIQLANGELKKVEDLKTEDFIQSAEISNDLKIDSSTVERIEDSH-----------   61
Rat_ATX1       SPAAASPTLPPYFMKGSIIQLANGELKKVEDLKTEDFIQSAEISNDLKIDSSTVERIEDSH-----------   61
Mouse_ATX1     SPTTASPTLPPYFMKGSIIQLANGELKKVEDLKTEDFIQSAEISNDLKIHSSTVERILESH-----------   61
Chicken_ATX1   SPATAPPTLPPYFMKGSIIQLANGELKKVEDLKTEDFIQSAEISNDLKIDSSTVERIEESH-----------   61
Frog_ATX1      SHMATAPTLPPYFMKGSIIQLANGELKKVEDLKTLDDFIQSADISNDLKIDSSTVERIKSSH-----------  61
Zebrafish_ATX1 PPPPSAPSMPPYFTKGSIIQLADGELKRVEDLKTEDFIQSAEISNELKIDSSTVERIDSSH-----------   61
Human_BOAT     PPPITSSHLPSHFMKGAIIQLATGELKRVEDLQTQDFVRSAEVSGGLKIDSSTVVDIQESQ-----------   61
Mouse_BOAT     PPPVTSSHLPSHFMKGAIIQLATGELKRVEDLQTQDFVRSAEVSGGLKIDSSTVVDIQESQ-----------   61
Zebrafish_BOAT QAPLLAPTGPSHFMKGAIIQLATGELKRVEDLQTQDFVRSAEMSGGLKIDSSMVVDIRASQQ----------   62
Dmelanog_ATX1  SSNGFSDDSASCFRRGSYIFLASGAMRRVEDIRTEDFIQSALRSQLFELREATVVRIDWSG-----------   61
Celegans_ATX1  GTIPASTYYPTHPMRGTQLNVANGNIKKVEDISSDDPLRCAAESDDVIVNASVIKSIKSTA-----------   61
Rat_HBP1       -----PSTVWHCFLKGTRLCFHKESKKEWQDV--EDFARAATCD-EEEIQMGTHKGYGSDGLELLSHEESVSF  65
Mouse_HBP1     -----PSTIWHCFLKGTRLCFHKESNKEWQDV--EDFARAASCDNEEKIQMGTHKGYGSDGLKLLSHKESVSF  66
Frog_HBP1      -----PQTVWHCFLKGTRLCFHKGRKKQWQDV--EDPAKSTRCRNERGIDSGTYKDVGSDGLKLVSYEBCVSY  66
Human_HBP1     -----PSTVWHCFLKGTRLCFHKGSNKEWQDV--EDFARAEGCDNEEDLQMGIHKGYGSDGLELLSHEESVSF  66
         .....210......220.......230.......240.......250.......260.......270...
```

```
         ....630........640.......650.......660.......670.......680.......690....
SS_Pred  ---EEEEEEE-------EEEEEEE----EEEE-----------EEEE-----EEEE--EEEE---------
Human_ATX1     SPGVAVIQFAVGEHRA-QVSVEVLVEYPFFVFGQGWSSCCPERTSQLFDLPCSKLSVGDVCISLTLK-----  127
Rat_ATX1       SPGVAVIQFAVGEHRA-QVSVEVLVEYPFFVFGQGWSSCCPERTSQLFDLPCSKLSVGDVCISLTLK-----  127
Mouse_ATX1     SPGVAVIQFAVGEHRA-QVSVEVLVEYPFFVFGQGWSSCCPERTSQLFDLPCSKLSVGDVCISLTLK-----  127
Chicken_ATX1   NPGIAVIQFAVGEHRA-QVSVEVLVEYPFFVFGQGWSSCCPERTSQLFDLPCSKLSVGDVCISLTLK-----  127
Frog_ATX1      SPGIAVVQFAVGEHRS-QVSVEVLIEYPFFVFGQGWSSCPERTSQMFDLPCSKLSVGDVCISLTLK-----  127
Zebrafish_ATX1 TANFAIIQFAVGAQRS-QVSVEVLVEYPFFVFGQGWSSCCPDRTTQLIELPCTKLSVGDVCISLTLK-----  127
Human_BOAT     WPGFVMLHFVVGEQQS-KVSIEVPPEHPFFVYGQGWSSCSPGRTTQLFSLPCHRLQVGDVCISISLQ-----  127
Mouse_BOAT     WPGFVMLHFVVGEQQS-KVSIEVPPEHPFFVYGQGWSSCSPGRTAQLFSLPCHRLQVGDVCISISLQ-----  127
Zebrafish_BOAT RPGLVALHFNVGEQQS-KVTIDVPPEHPFFVYGQGWSSCSPERTAQLYGLTCHHLQVGDVCVSVTLQ-----  128
Dmelanog_ATX1  CPSLVTLTFSYDTHHA-KMDLQVQPGHPMFVYGQGWASCDPQLSLQLYELKGQQLQVGDICLSLVPN-----  127
Celegans_ATX1  --GSVTIIFETGIEKQ-LIPLKCQVEHPFFVLGKGWCSCNPRKSGENYGLDCEILQVIDVCIVLTRN-----  125
Rat_HBP1       GESVLKLTFDPGTVEDGLLTVECKLDHPFYVKNKGWSSFYPSLTVVQHGIPCCEIHIGDVCLPPGHPCAINF  137
Mouse_HBP1     GESVLKLTFDPGTVEDGLLTVECKLDHPFYVKNKGWSSFYPSLTVVQHGIPCCEIHIGDVCLPPGHPCAINF  138
Frog_HBP1      GKSVLQLTFDPGTTVGLLTVCKLDHPYYVKNKGWSSFYPSLTVVQHGIPCCKIHVGDVCLPPGHPCAINF  138
Human_HBP1     GESVLKLTFDPGTXEDGLLTVECKLDHPFYVKNKVWSSFYPSLTVVQHGIPCCEVHIGNVCLPPGHPCAINF  138
         ....280.......290.......300.......310.......320.......330.......340.....
```

Fig. 4. Sequence alignment of the AXH from ataxin-1 (ATX1), ataxin-1 paralogue BOAT (Brother Of ATaxin-1) and HBP1 from different species. The alignment was prepared by ClustalX (version 2) (Larkin et al., 2007) and is based exclusively on sequence similarity. The secondary structure of the ataxin-1 AXH domain as predicted by Jpred 3 (Cole et al., 2008) is shown on the top for reference.
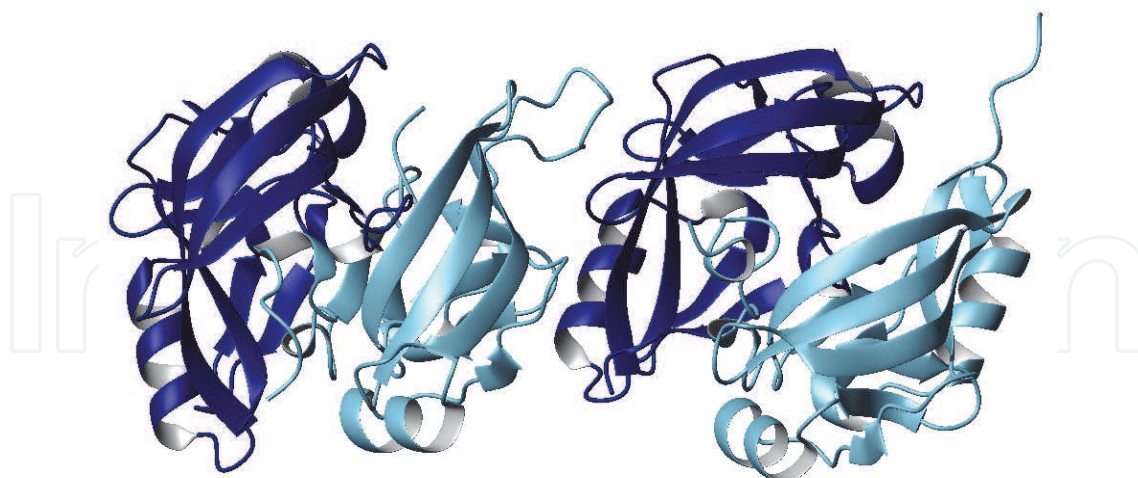

Fig. 5. X-ray structure of the AXH domain of ataxin-1 (PDB entry 1OA8) (Chen et al., 2004). The monomers forming the dimer of dimers observed in the structure are alternatively indicated in dark and light blue. Detailed analysis shows that they are not related by symmetry.

An even bigger surprise was revealed by structure determination of the AXH domain of HBP1. Whilst on the pure basis of sequence homology it would have been reasonable to

assume that the structure of each of the two domains could be easily modelled from the other used as a template, an unexpected result came out when the solution structure of the HBP1 AXH domain was solved by NMR spectroscopy (de Chiara et al., 2005a) (**Figures 6 and 7**).
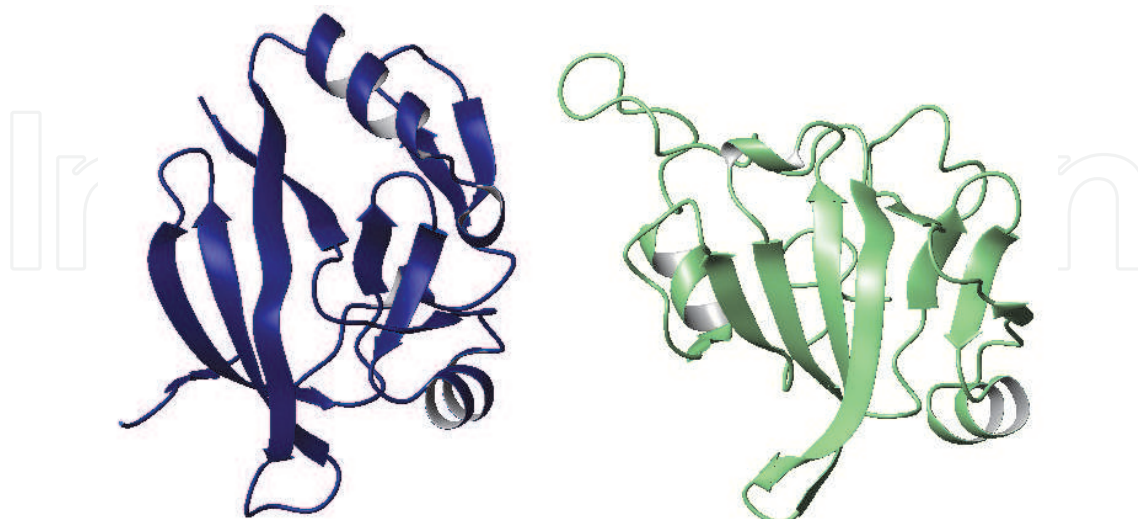

Fig. 6. Comparison between the structure of the AXH domains of ataxin-1 (monomer A) (left) (PDB entry 1OA8) (Chen et al., 2004) and HBP1 (right) (PDB entry 1V06) (de Chiara et al., 2005a). The N-terminal of the two monomers show the same elements of secondary structure arranged in a different topology.

Fig. 7. Alignment based on structural comparison between the AXH from taxin-1 and HBP1 (Chen et al., 2004; de Chiara et al., 2005a). SS_ATX1 and SS_HBP1:experimental secondary structure of ataxin-1 and HBP1.

When comparing the structure of the HBP1 domain, which is monomeric (de Chiara et al., 2003; de Chiara et al., 2005a), to any of the ataxin-1 X-ray monomers, only the C-terminal part, representing the core of the OB-fold, is structurally superposable, while the N-terminus adopts a different topology, despite maintaining the same secondary structure elements along the sequence (**Figure 6**). Only a structure-based comparison allowed us to realign the sequences and to correctly position the HBP1 long-loop insertion, which was originally set between beta-3 and beta-4, between helix-1 and beta-3 (**Figure 7**). These findings support the possibility that the AXH motif is intrinsically able to adopt different topologies.

6.1.3.2 The function of the ataxin-1 AXH domain

Further studies on the role of the OB-fold of the AXH domain have shown that this region is designed to mediate interactions both with nucleic acids and proteins. The crystal structure of the ataxin-1 AXH allowed us to rationalize previous literature notions on the ability of the AXH to bind RNA homopolimers in vitro with the same nucleotide preference as full-length ataxin-1 (de Chiara et al., 2003; Yue et al., 2001). In addition to a direct binding to RNA through the AXH domain, the protein was found to co-localize with RNA also when the AXH domain was deleted, thus suggesting an involvement of other RNA binding proteins in the ataxin-1 interactome (de Chiara et al., 2005b). Recent findings on the ability of ataxin-1 to interact with splicing factors through a short motif localized C-terminally to the AXH domain opened the intriguing possibility that the protein may be involved in pre-mRNA processing at the level of the splicing machinery (de Chiara et al., 2009; Lim et al., 2008). However, no RNA targets have, as yet, been identified and more research is needed to address the question of whether the protein may play a role in RNA metabolism and/or nuclear RNA export as suggested also by co-localization with the mRNA export factor TAP/NXF1 (Irwin et al., 2005).

As for the ability of the AXH domain to mediate protein-protein interactions, several binding partners with transcriptional activity have been identified whose interaction with ataxin-1 is abolished when the AXH domain is deleted: the silencing mediator of retinoid and thyroid hormone receptors SMRT/SMRTER (Tsai et al., 2004), the repressor Capicua (Lam et al., 2006), the transcription factors Sensless/Gfi-1 (Tsuda et al., 2005) and Sp1 (Goold et al., 2007). A potential role for ataxin-1 in transcriptional regulation was suggested at a very early stage by the homology with HBP1 (Sampson et al., 2001; Tevosian et al., 1997). A general read-out assay for repression of transcription (de Chiara et al., 2005b) confirmed that the AXH domain represses transcription when tethered to DNA similarly to what was observed for full-length ataxin-1 (Tsai et al., 2004). However, cross-linking experiments showed that there is no direct binding between DNA and AXH domain indicating that the interaction is mediated by other co-transcriptional regulators (de Chiara et al., 2005b), as also confirmed experimentally later on (Bolger et al., 2007; Goold et al., 2007; Lam et al., 2006; Serra et al., 2006; Tsuda et al., 2005).

6.1.4 The C-terminal IUR

The region downstream to the AXH domain up to the C-terminal end of the protein (amino acids 690-816) represents an example of possible conflict between the results of different predictors. Disopred2, PONDR, IUPRED and DisEMBL predict the C-terminus as an almost completely disordered region **(Figures 3)**. According to GlobPlot (http://globplot.embl.de), which is based on the Russell/Linding scale (Linding et al., 2003b), the region 703-786 is

predicted as a potential globular domain. There is also no agreement with the prediction from the SMART server which, instead, identifies the AXH as the only folded region in the protein (**Figures 2**). While still awaiting for a systematic experimental validation, we can already comment on these results in light of our findings.

6.1.4.1 Prediction of ELMs in the C-terminal IUR: A three-way molecular switch in ataxin-1 C-terminus

Consistent with the presence of disorder, several linear motifs were predicted in the protein C-terminus. Among these three overlapping linear motifs identified downstream to AXH were experimentally verified: a nuclear localization signal (NLS) (771-774) (Klement et al., 1998), a 14-3-3 binding motif (774-778: key conservation RxxSxP) (Chen et al., 2003) and a UHM ligand motif (ULM) (771-776) (de Chiara et al., 2009), present in proteins associated with splicing. These motifs represent a three-way molecular switch which plays an important role both for the function of the native protein and for pathogenesis. In addition to the expansion of the polyQ tract, nuclear localization is a strict requirement for the development of the pathology. Expanded ataxin-1 with mutated NLS fails to enter the nucleus and does not cause aggregation that is the typical phenotypic hallmark of the SCA1 pathology (Klement et al., 1998). Further to polyQ expansion and nuclear localization, phosphorylation of S776 has been identified as a condition necessary for development of SCA1 (Emamian et al., 2003). Phosphorylation of S776 has been confirmed to occur in vivo (Emamian et al., 2003; Jorgensen et al., 2009) and is required for recognition of ataxin-1 by the protein 14-3-3, a molecular adaptor which modulates, in a phosphorylation-dependent manner, the function of different proteins in their specific context (Chen et al., 2003). Mutation of S776 to an alanine in expanded ataxin-1, despite not affecting nuclear localization, prevented the development of the SCA1 phenotype (Emamian et al., 2003).

Recently, an UHM ligand motif (ULM) predicted by the ELM server in the C-terminus of the ataxin-1 sequence has been experimentally validated and characterized (de Chiara et al., 2009). The ULM motif was first identified in the splicing factors SF1 and SAP155 and shown to bind the UHM domain of U2AF65 (Corsini et al., 2007). Ataxin-1 ULM (771-776) strongly overlaps with the 14-3-3 ligand motif (774-778). However, whilst phosphorylation of S776 is crucial for recognition by 14-3-3, it only marginally affects the interaction with U2AF65, increasing the dissociation constant by only ~3 folds. Being the K_d between ataxin-1 S776-phosphorylated ULM and 14-3-3 (ζ isoform) two orders of magnitude smaller than U2AF65 (0.4 μM versus 36 μM) it was possible to conclude that, when S776 is phosphorylated, 14-3-3 is able to displace U2AF65. Under these conditions, the 14-3-3-bound expanded ataxin-1 is prone to aggregation. The S776A mutation, which hampers the interaction with 14-3-3, still allows the interaction with UHM domain of U2AF65 and potentially other splicing factors. These interactions, likely because of the extended dimension of the spliceosome complex, may play a protective role and prevent aggregation. Our findings allowed us to conclude that phosphorylation of S776 provides the switch that regulates binding of ataxin-1 to the protein 14-3-3 and components of the spliceosome, and suggests that pathology develops when aggregation competes with native interactions (**Figure 8**). This example also shows how the investigation of the native function of the polyQ proteins have provided valuable hints for understanding the molecular mechanisms of pathogenesis.
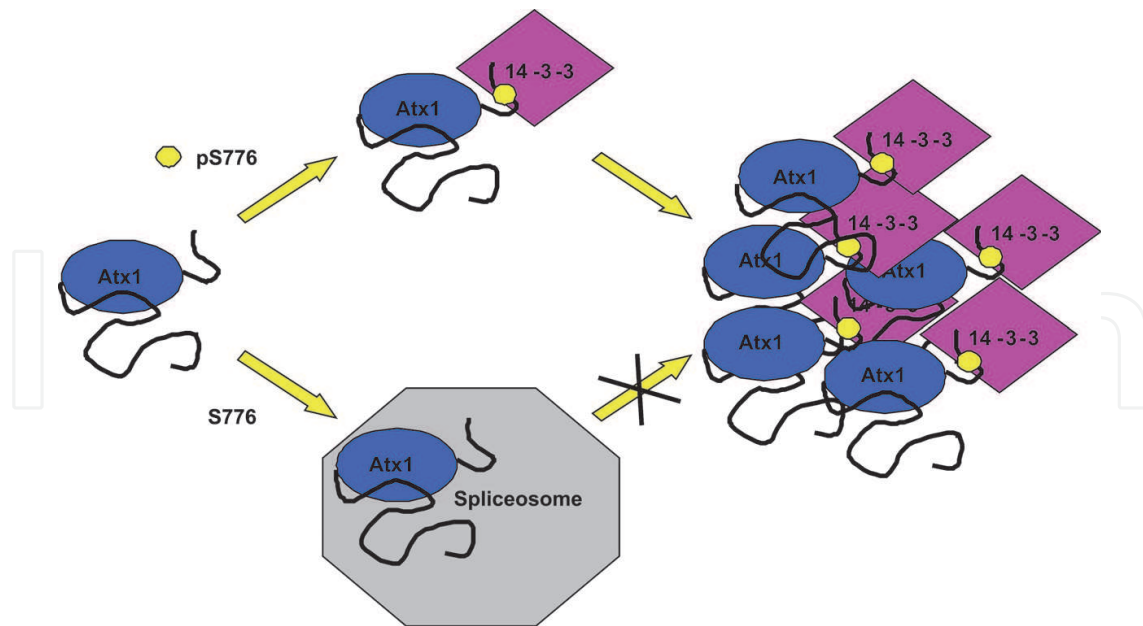
Fig. 8. A model of the role played by phosphorylation of S776 in the modulation of expanded ataxin-1 aggregation.

7. Conclusions

We have discussed here the concept of IUPs and IURs and shown how their detailed bioinformatics analysis can assist structural and functional studies using ataxin-1 as a paradigmatic example. Ataxin-1, like other members of the polyQ pathology family, is mostly composed of IUPs. Very little is still known about this protein despite its involvement in human neurodegeneration, yet this knowledge is essential for designing specific therapeutic interventions. Identification of both structured (the AXH domain) and unstructured linear functional motifs has played a key role in advancing our knowledge in the ataxin-1 function in the cell. More advanced information will undoubtedly come from experimental analysis of long stretches of the protein if the formidable challenges of their recombinant production in a pure and stable form can be circumvented. It is also evident from the example reported here how new approaches in the identification and study of IUPs might be highly helpful to advance the field.

8. References

Banfi, S., Servadio, A., Chung, M.Y., Kwiatkowski, T.J., Jr., McCall, A.E., Duvick, L.A., Shen, Y., Roth, E.J., Orr, H.T., & Zoghbi, H.Y. (1994). Identification and Characterization of the Gene Causing Type 1 Spinocerebellar Ataxia. *Nature Genetics*, Vol. 7, No. 4, (August 1994), pp. 513-520, ISSN 1061-4036

Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., & Svergun, D.I. (2007). Structural Characterization of Flexible Proteins Using Small-Angle X-Ray Scattering. *Journal of the American Chemical Society*, Vol. 129, No. 17, (May 2007), pp. 5656-5664, ISSN 0002-7863

Boesch, C., Bundi, A., Oppliger, M., & Wuthrich, K. (1978). 1h Nuclear-Magnetic-Resonance Studies of the Molecular Conformation of Monomeric Glucagon in Aqueous

Solution. *European Journal of Biochemistry,* Vol. 91, No. 1, (November 1978), pp. 209-214, ISSN 0014-2956

Bolger, T.A., Zhao, X., Cohen, T.J., Tsai, C.C., &Yao, T.P. (2007). The Neurodegenerative Disease Protein Ataxin-1 Antagonizes the Neuronal Survival Function of Myocyte Enhancer Factor-2. *Journal of Biologic Chemistry,* Vol. 282, No. 40, (October 2007), pp. 29186-29192, ISSN 0021-9258

Bonvin, A.M., & Brunger, A.T. (1996). Do Noe Distances Contain Enough Information to Assess the Relative Populations of Multi-Conformer Structures? *Journal of Biomolecular NMR,* Vol. 7, No. 1, (January 1996), pp. 72-76, ISSN 0925-2738

Burright, E.N., Davidson, J.D., Duvick, L.A., Koshy, B., Zoghbi, H.Y., & Orr, H.T. (1997). Identification of a Self-Association Region within the Sca1 Gene Product, Ataxin-1. *Human Molecular Genetics,* Vol. 6, No. 4, (April 1997), pp. 513-518, ISSN 0964-6906

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., & Zoghbi, H.Y. (2003). Interaction of Akt-Phosphorylated Ataxin-1 with 14-3-3 Mediates Neurodegeneration in Spinocerebellar Ataxia Type 1. *Cell,* Vol. 113, No. 4, (May 2003), pp. 457-468, ISSN 0092-8674

Chen, Y.W., Allen, M.D., Veprintsev, D.B., Lowe, J., & Bycroft, M. (2004). The Structure of the Axh Domain of Spinocerebellar Ataxin-1. *Journal of Biologic Chemistry,* Vol. 279, No. 5, (January 2004), pp. 3758-3765, ISSN 0021-9258

Clore, G.M., & Schwieters, C.D. (2004a). How Much Backbone Motion in Ubiquitin Is Required to Account for Dipolar Coupling Data Measured in Multiple Alignment Media as Assessed by Independent Cross-Validation? *Journal of the American Chemical Society,* Vol. 126, No. 9, (March 2004a), pp. 2923-2938, ISSN 0002-7863

Clore, G.M., & Schwieters, C.D. (2004b). Amplitudes of Protein Backbone Dynamics and Correlated Motions in a Small Alpha/Beta Protein: Correspondence of Dipolar Coupling and Heteronuclear Relaxation Measurements. *Biochemistry,* Vol. 43, No. 33, (Aug 24 2004b), pp. 10678-10691, ISSN 0006-2960

Cole, C., Barber, J.D., & Barton, G.J. (2008). The Jpred 3 Secondary Structure Prediction Server. *Nucleic Acids Research,* Vol. 36, Web Server issue, (July 2008), pp. W197-201, ISSN 1362-4962

Corsini, L., Bonnal, S., Basquin, J., Hothorn, M., Scheffzek, K., Valcarcel, J., & Sattler, M. (2007). U2af-Homology Motif Interactions Are Required for Alternative Splicing Regulation by Spf45. *Nature Structure and Molecular Biology,* Vol. 14, No. 7, (July 2007), pp. 620-629, ISSN 1545-9993

Csizmok, V., Dosztanyi, Z., Simon, I., & Tompa, P. (2007). Towards Proteomic Approaches for the Identification of Structural Disorder. *Current Protein & Peptide Science,* Vol. 8, No. 2, (April 2007), pp. 173-179, ISSN 1389-2037

de Chiara, C., Giannini, C., Adinolfi, S., de Boer, J., Guida, S., Ramos, A., Jodice, C., Kioussis, D., & Pastore, A. (2003). The Axh Module: An Independently Folded Domain Common to Ataxin-1 and Hbp1. *FEBS Letters,* Vol. 551, No. 1-3, (September 2003), pp. 107-112, ISSN 0014-5793

de Chiara, C., Menon, R.P., Adinolfi, S., de Boer, J., Ktistaki, E., Kelly, G., Calder, L., Kioussis, D., & Pastore, A. (2005a). The Axh Domain Adopts Alternative Folds the Solution Structure of Hbp1 Axh. *Structure,* Vol. 13, No. 5, (May 2005a), pp. 743-753, ISSN 0969-2126

de Chiara, C., Menon, R.P., Dal Piaz, F., Calder, L., &Pastore, A. (2005b). Polyglutamine Is Not All: The Functional Role of the Axh Domain in the Ataxin-1 Protein. *Journal of Molecular Biology,* Vol. 354, No. 4, (December 2005b), pp. 883-893, ISSN 0022-2836

de Chiara, C., Menon, R.P., Strom, M., Gibson, T.J., & Pastore, A. (2009). Phosphorylation of S776 and 14-3-3 Binding Modulate Ataxin-1 Interaction with Splicing Factors. *PLoS One,* Vol. 4, No. 12, (December 2009), pp. e8372, ISSN 1932-6203

Dephoure, N., Zhou, C., Villen, J., Beausoleil, S.A., Bakalarski, C.E., Elledge, S.J., & Gygi, S.P. (2008). A Quantitative Atlas of Mitotic Phosphorylation. *Proceedings of the National Academy of Sciences U S A,* Vol. 105, No. 31, (August 2008), pp. 10762-10767, ISSN 1091-6490

Diella, F., Haslam, N., Chica, C., Budd, A., Michael, S., Brown, N.P., Trave, G., & Gibson, T.J. (2008). Understanding Eukaryotic Linear Motifs and Their Role in Cell Signaling and Regulation. *Frontiers in Bioscience,* Vol. 13, No. (May 2008), pp. 6580-6603, ISSN 1093-4715

Dosztanyi, Z., Csizmok, V., Tompa, P., & Simon, I. (2005a). The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins. *Journal of Molecular Biology,* Vol. 347, No. 4, (April 2005a), pp. 827-839, ISSN 0022-2836

Dosztanyi, Z., Csizmok, V., Tompa, P., & Simon, I. (2005b). Iupred: Web Server for the Prediction of Intrinsically Unstructured Regions of Proteins Based on Estimated Energy Content. *Bioinformatics,* Vol. 21, No. 16, (August 2005b), pp. 3433-3434, ISSN 1367-4803

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., & Obradovic, Z. (2001). Intrinsically Disordered Protein. *Journal of Molecular Graphics and Modelling,* Vol. 19, No. 1, pp. 26-59, ISSN 1093-3263

Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., & Obradovic, Z. (2002). Intrinsic Disorder and Protein Function. *Biochemistry,* Vol. 41, No. 21, (May 2002), pp. 6573-6582, ISSN 0006-2960

Emamian, E.S., Kaytor, M.D., Duvick, L.A., Zu, T., Tousey, S.K., Zoghbi, H.Y., Clark, H.B., & Orr, H.T. (2003). Serine 776 of Ataxin-1 Is Critical for Polyglutamine-Induced Disease in Sca1 Transgenic Mice. *Neuron,* Vol. 38, No. 3, (May 2003), pp. 375-387, ISSN 0896-6273

Evans, P.R., & Owen, D.J. (2002). Endocytosis and Vesicle Trafficking. *Current Opinion in Structural Biology,* Vol. 12, No. 6, (December 2002), pp. 814-821, ISSN 0959-440X

Fennen, J., Torda, A.E., & van Gunsteren, W.F. (1995). Structure Refinement with Molecular Dynamics and a Boltzmann-Weighted Ensemble. *Journal of Biomolecular NMR,* Vol. 6, No. 2, (September 1995), pp. 163-170, ISSN 0925-2738

Ferron, F., Longhi, S., Canard, B., & Karlin, D. (2006). A Practical Overview of Protein Disorder Prediction Methods. *Proteins,* Vol. 65, No. 1, (October 2006), pp. 1-14, ISSN 1097-0134

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E.L., Eddy, S.R., & Bateman, A. (2010). The Pfam Protein Families Database. *Nucleic Acids Research,* Vol. 38, Database issue, (January 2010), pp. D211-222, ISSN 1362-4962

Genis, D., Matilla, T., Volpini, V., Rosell, J., Davalos, A., Ferrer, I., Molins, A., & Estivill, X. (1995). Clinical, Neuropathologic, and Genetic Studies of a Large Spinocerebellar Ataxia Type 1 (Sca1) Kindred: (Cag)N Expansion and Early Premonitory Signs and Symptoms. *Neurology,* Vol. 45, No. 1, (January 1995), pp. 24-30, ISSN 0028-3878

Gibson, T.J. (2009). Cell Regulation: Determined to Signal Discrete Cooperation. *Trends in Biochemical Sciences,* Vol. 34, No. 10, (October 2009), pp. 471-482, ISSN 0968-0004

Goold, R., Hubank, M., Hunt, A., Holton, J., Menon, R.P., Revesz, T., Pandolfo, M., & Matilla-Duenas, A. (2007). Down-Regulation of the Dopamine Receptor D2 in Mice Lacking Ataxin 1. *Human Molecular Genetics,* Vol. 16, No. 17, (September 2007), pp. 2122-2134, ISSN 0964-6906

Greaser, M.L., Wang, S.M., Berri, M., Mozdziak, P., & Kumazawa, Y. (2000). Sequence and Mechanical Implications of Titin's Pevk Region. *Advances in Experimental Medicine and Biology,* Vol. 481, pp. 53-63; discussion 64-56, 107-110, ISSN 0065-2598

Iakoucheva, L.M., Brown, C.J., Lawson, J.D., Obradovic, Z., & Dunker, A.K. (2002). Intrinsic Disorder in Cell-Signaling and Cancer-Associated Proteins. *Journal of Molecular Biology,* Vol. 323, No. 3, (October 2002), pp. 573-584, ISSN 0022-2836

Irwin, S., Vandelft, M., Pinchev, D., Howell, J.L., Graczyk, J., Orr, H.T., & Truant, R. (2005). Rna Association and Nucleocytoplasmic Shuttling by Ataxin-1. *Journal of Cell Science,* Vol. 118, No. Pt 1, (January 2005), pp. 233-242, ISSN 0021-9533

Jayaraman, M., Kodali, R., & Wetzel, R. (2009). The Impact of Ataxin-1-Like Histidine Insertions on Polyglutamine Aggregation. *Protein Engineering Design Selection,* Vol. 22, No. 8, (August 2009), pp. 469-478, ISSN 1741-0134

Jorgensen, N.D., Andresen, J.M., Lagalwar, S., Armstrong, B., Stevens, S., Byam, C.E., Duvick, L.A., Lai, S., Jafar-Nejad, P., Zoghbi, H.Y., Clark, H.B., & Orr, H.T. (2009). Phosphorylation of Atxn1 at Ser776 in the Cerebellum. *Journal of Neurochemistry,* Vol. 110, No. 2, (July 2009), pp. 675-686, ISSN 1471-4159

Klement, I.A., Skinner, P.J., Kaytor, M.D., Yi, H., Hersch, S.M., Clark, H.B., Zoghbi, H.Y., & Orr, H.T. (1998). Ataxin-1 Nuclear Localization and Aggregation: Role in Polyglutamine-Induced Disease in Sca1 Transgenic Mice. *Cell,* Vol. 95, No. 1, (October 1998), pp. 41-53, ISSN 0092-8674

Konagurthu, A.S., &Lesk, A.M. Cataloging Topologies of Protein Folding Patterns. *J Mol Recognit,* Vol. 23, No. 2, (March-April 2010) pp. 253-257, ISSN 1099-1352

Kuszewski, J., Gronenborn, A.M., & Clore, G.M. (1996). Improving the Quality of Nmr and Crystallographic Protein Structures by Means of a Conformational Database Potential Derived from Structure Databases. *Protein Science,* Vol. 5, No. 6, (June 1996), pp. 1067-1080, ISSN 0961-8368

Labeit, S., &Kolmerer, B. (1995). Titins: Giant Proteins in Charge of Muscle Ultrastructure and Elasticity. *Science,* Vol. 270, No. 5234, (October 1995), pp. 293-296, ISSN 0036-8075

Lam, Y.C., Bowman, A.B., Jafar-Nejad, P., Lim, J., Richman, R., Fryer, J.D., Hyun, E.D., Duvick, L.A., Orr, H.T., Botas, J., & Zoghbi, H.Y. (2006). Ataxin-1 Interacts with the Repressor Capicua in Its Native Complex to Cause Sca1 Neuropathology. *Cell,* Vol. 127, No. 7, (December 2006), pp. 1335-1347, ISSN 0092-8674

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., &

Higgins, D.G. (2007). Clustal W and Clustal X Version 2.0. *Bioinformatics,* Vol. 23, No. 21, (November 2007), pp. 2947-2948, ISSN 1367-4811

Letunic, I., Doerks, T., & Bork, P. (2009). Smart 6: Recent Updates and New Developments. *Nucleic Acids Research,* Vol. 37, Database issue, (January 2009), pp. D229-232, ISSN 1362-4962

Lim, J., Crespo-Barreto, J., Jafar-Nejad, P., Bowman, A.B., Richman, R., Hill, D.E., Orr, H.T., & Zoghbi, H.Y. (2008). Opposing Effects of Polyglutamine Expansion on Native Protein Complexes Contribute to Sca1. *Nature,* Vol. 452, No. 7188, (April 2008), pp. 713-718, ISSN 1476-4687

Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., & Russell, R.B. (2003a). Protein Disorder Prediction: Implications for Structural Proteomics. *Structure,* Vol. 11, No. 11, (November 2003a), pp. 1453-1459, ISSN 0969-2126

Linding, R., Russell, R.B., Neduva, V., & Gibson, T.J. (2003b). Globplot: Exploring Protein Sequences for Globularity and Disorder. *Nucleic Acids Research,* Vol. 31, No. 13, (July 2003b), pp. 3701-3708, ISSN 1362-4962

Masino, L., Kelly, G., Leonard, K., Trottier, Y., & Pastore, A. (2002). Solution Structure of Polyglutamine Tracts in Gst-Polyglutamine Fusion Proteins. *FEBS Letters,* Vol. 513, No. 2-3, (February 2002), pp. 267-272, ISSN 0014-5793

Mittag, T., & Forman-Kay, J.D. (2007). Atomic-Level Characterization of Disordered Protein Ensembles. *Current Opinion in Structural Biology,* Vol. 17, No. 1, (February 2007), pp. 3-14, ISSN 0959-440X

Murzin, A.G. (1993). Ob(Oligonucleotide/Oligosaccharide Binding)-Fold: Common Structural and Functional Solution for Non-Homologous Sequences. *The EMBO Journal,* Vol. 12, No. 3, (March 1993), pp. 861-867, ISSN 0261-4189

Mushegian, A.R., Bassett, D.E., Jr., Boguski, M.S., Bork, P., & Koonin, E.V. (1997). Positionally Cloned Human Disease Genes: Patterns of Evolutionary Conservation and Functional Motifs. *Proceedings of the National Academy of Sciences U.S.A.,* Vol. 94, No. 11, (May 1997), pp. 5831-5836, ISSN 0027-8424

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., & Kanazawa, I. (2002). Interaction between Mutant Ataxin-1 and Pqbp-1 Affects Transcription and Cell Death. *Neuron,* Vol. 34, No. 5, (May 2002), pp. 701-713, ISSN 0896-6273

Orr, H.T., Chung, M.Y., Banfi, S., Kwiatkowski, T.J., Jr., Servadio, A., Beaudet, A.L., McCall, A.E., Duvick, L.A., Ranum, L.P., & Zoghbi, H.Y. (1993). Expansion of an Unstable Trinucleotide Cag Repeat in Spinocerebellar Ataxia Type 1. *Nature Genetics,* Vol. 4, No. 3, (July 1993), pp. 221-226, ISSN 1061-4036

Orr, H.T., & Zoghbi, H.Y. (2007). Trinucleotide Repeat Disorders. *Annual Review in Neuroscience,* Vol. 30, pp. 575-621, ISSN 0147-006X

Puntervoll, P., Linding, R., Gemund, C., Chabanis-Davidson, S., Mattingsdal, M., Cameron, S., Martin, D.M., Ausiello, G., Brannetti, B., Costantini, A., Ferre, F., Maselli, V., Via, A., Cesareni, G., Diella, F., Superti-Furga, G., Wyrwicz, L., Ramu, C., McGuigan, C., Gudavalli, R., Letunic, I., Bork, P., Rychlewski, L., Kuster, B., Helmer-Citterich, M., Hunter, W.N., Aasland, R., & Gibson, T.J. (2003). Elm Server: A New Resource for Investigating Short Functional Sites in Modular Eukaryotic Proteins. *Nucleic Acids Research,* Vol. 31, No. 13, (July 2003), pp. 3625-3630, ISSN 1362-4962

Radivojac, P., Obradovic, Z., Brown, C.J., & Dunker, A.K. (2003). Prediction of Boundaries between Intrinsically Ordered and Disordered Protein Regions. *Pacific Symposium on Biocomputing,* Vol. 8, pp. 216-227, ISSN 1793-5091

Radivojac, P., Iakoucheva, L.M., Oldfield, C.J., Obradovic, Z., Uversky, V.N., & Dunker, A.K. (2007). Intrinsic Disorder and Functional Proteomics. *Biophysical Journal,* Vol. 92, No. 5, (March 2007), pp. 1439-1456, ISSN 1542-0086

Romero, P., Pbradovic, Z., Kissinger, C.R., Villafranca, J.E., & Dunker, A.K. Identifying Disordered Regions in Proteins from Amino Acid Sequences; 1997; Houston, TX. pp. 90-95.

Sampson, E.M., Haque, Z.K., Ku, M.C., Tevosian, S.G., Albanese, C., Pestell, R.G., Paulson, K.E., & Yee, A.S. (2001). Negative Regulation of the Wnt-Beta-Catenin Pathway by the Transcriptional Repressor Hbp1. *The Embo Journal,* Vol. 20, No. 16, (August 2001), pp. 4500-4511, ISSN 0261-4189

Schaeffer, R.D., &Daggett, V. Protein Folds and Protein Folding. *Protein Engineering Design and Selection,* Vol. 24, No. 1-2, (January 2010), pp. 11-19, ISSN 1741-0134

Schultz, J., Milpetz, F., Bork, P., & Ponting, C.P. (1998). Smart, a Simple Modular Architecture Research Tool: Identification of Signaling Domains. *Proceedings of the National Academy of Sciences U S A,* Vol. 95, No. 11, (May 1998), pp. 5857-5864, ISSN 0027-8424

Serra, H.G., Duvick, L., Zu, T., Carlson, K., Stevens, S., Jorgensen, N., Lysholm, A., Burright, E., Zoghbi, H.Y., Clark, H.B., Andresen, J.M., &Orr, H.T. (2006). Roralpha-Mediated Purkinje Cell Development Determines Disease Severity in Adult Sca1 Mice. *Cell,* Vol. 127, No. 4, (November 2006), pp. 697-708, ISSN 0092-8674

Shoemaker, B.A., Portman, J.J., & Wolynes, P.G. (2000). Speeding Molecular Recognition by Using the Folding Funnel: The Fly-Casting Mechanism. *Proceedings of the National Academy of Sciences U S A,* Vol. 97, No. 16, (August 2000), pp. 8868-8873, ISSN 0027-8424

Sjekloća, L., Pauwels, K., & Pastore, A. (2011) *FEBS Journal*, Accepted Article; doi: 10.1111/j.1742-4658.2011.08108.x

Tevosian, S.G., Shih, H.H., Mendelson, K.G., Sheppard, K.A., Paulson, K.E., & Yee, A.S. (1997). Hbp1: A Hmg Box Transcriptional Repressor That Is Targeted by the Retinoblastoma Family. *Genes & Development,* Vol. 11, No. 3, (February 1997), pp. 383-396, ISSN 0890-9369

Tompa, P., & Csermely, P. (2004). The Role of Structural Disorder in the Function of Rna and Protein Chaperones. *Faseb Journal,* Vol. 18, No. 11, (August 2004), pp. 1169-1175, ISSN 1530-6860

Tompa, P. (2005). The Interplay between Structure and Function in Intrinsically Unstructured Proteins. *FEBS Letters,* Vol. 579, No. 15, (June 2005), pp. 3346-3354, ISSN 0014-5793

Tompa, P., & Fuxreiter, M. (2008). Fuzzy Complexes: Polymorphism and Structural Disorder in Protein-Protein Interactions. *Trends in Biochemical Sciences,* Vol. 33, No. 1, (January 2008), pp. 2-8, ISSN 0968-0004

Tsai, C.C., Kao, H.Y., Mitzutani, A., Banayo, E., Rajan, H., McKeown, M., & Evans, R.M. (2004). Ataxin 1, a Sca1 Neurodegenerative Disorder Protein, Is Functionally Linked to the Silencing Mediator of Retinoid and Thyroid Hormone Receptors.

Proceedings of the National Academy of Sciences U.S.A., Vol. 101, No. 12, (March 2004), pp. 4047-4052, ISSN 0027-8424

Tsuda, H., Jafar-Nejad, H., Patel, A.J., Sun, Y., Chen, H.K., Rose, M.F., Venken, K.J., Botas, J., Orr, H.T., Bellen, H.J., & Zoghbi, H.Y. (2005). The Axh Domain of Ataxin-1 Mediates Neurodegeneration through Its Interaction with Gfi-1/Senseless Proteins. *Cell,* Vol. 122, No. 4, (August 2005), pp. 633-644, ISSN 0092-8674

Uversky, V.N. (2002). What Does It Mean to Be Natively Unfolded? *European Journal of Biochemistry,* Vol. 269, No. 1, (January 2002), pp. 2-12, ISSN 0014-2956

Vendruscolo, M., & Paci, E. (2003). Protein Folding: Bringing Theory and Experiment Closer Together. *Current Opinion in Structural Biology,* Vol. 13, No. 1, (February 2003), pp. 82-87, ISSN 0959-440X

Vendruscolo, M., & Dobson, C.M. (2005). Towards Complete Descriptions of the Free-Energy Landscapes of Proteins. *Philos Transact A Math Phys Eng Sci,* Vol. 363, No. 1827, (February 2005), pp. 433-450; discussion 450-432, ISSN 1364-503X

Vierra-Green, C.A., Orr, H.T., Zoghbi, H.Y., & Ferrington, D.A. (2005). Identification of a Novel Phosphorylation Site in Ataxin-1. *Biochimica et Biophysica Acta,* Vol. 1744, No. 1, (May 2005), pp. 11-18, ISSN 0006-3002

Ward, J.J., McGuffin, L.J., Bryson, K., Buxton, B.F., & Jones, D.T. (2004a). The Disopred Server for the Prediction of Protein Disorder. *Bioinformatics,* Vol. 20, No. 13, (September 2004a), pp. 2138-2139, ISSN 1367-4803

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., & Jones, D.T. (2004b). Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *Journal of Molecular Biology,* Vol. 337, No. 3, (March 2004b), pp. 635-645, ISSN 0022-2836

Wright, P.E., & Dyson, H.J. (1999). Intrinsically Unstructured Proteins: Re-Assessing the Protein Structure-Function Paradigm. *Journal of Molecular Biology,* Vol. 293, No. 2, (October 1999), pp. 321-331, ISSN 0022-2836

Wright, P.E., & Dyson, H.J. (2009). Linking Folding and Binding. *Current Opinion in Structural Biology,* Vol. 19, No. 1, (February 2009), pp. 31-38, ISSN 1879-033X

Yue, S., Serra, H.G., Zoghbi, H.Y., & Orr, H.T. (2001). The Spinocerebellar Ataxia Type 1 Protein, Ataxin-1, Has Rna-Binding Activity That Is Inversely Affected by the Length of Its Polyglutamine Tract. *Human Molecular Genetics,* Vol. 10, No. 1, (January 2001), pp. 25-30, ISSN 0964-6906

Zoghbi, H.Y., & Orr, H.T. (2008). Pathogenic Mechanisms of a Polyglutamine Mediated Neurodegenerative Disease: Sca1. *Journal of Biologic Chemistry,* Vol. 284, No. 12, (October 2008), pp. 7425-7429, ISSN 0021-9258
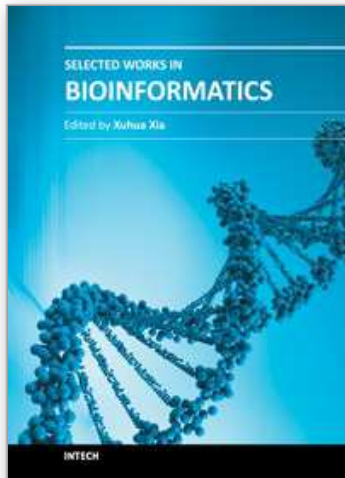
Selected Works in Bioinformatics

Edited by Dr. Xuhua Xia

This book consists of nine chapters covering a variety of bioinformatics subjects, ranging from database resources for protein allergens, unravelling genetic determinants of complex disorders, characterization and prediction of regulatory motifs, computational methods for identifying the best classifiers and key disease genes in large-scale transcriptomic and proteomic experiments, functional characterization of inherently unfolded proteins/regions, protein interaction networks and flexible protein-protein docking. The computational algorithms are in general presented in a way that is accessible to advanced undergraduate students, graduate students and researchers in molecular biology and genetics. The book should also serve as stepping stones for mathematicians, biostatisticians, and computational scientists to cross their academic boundaries into the dynamic and ever-expanding field of bioinformatics.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Cesira de Chiara and Annalisa Pastore (2011). Prediction and Experimental Detection of Structural and Functional Motifs in Intrinsically Unfolded Proteins, Selected Works in Bioinformatics, Dr. Xuhua Xia (Ed.), ISBN: 978-953-307-281-4, InTech, Available from: http://www.intechopen.com/books/selected-works-in-bioinformatics/prediction-and-experimental-detection-of-structural-and-functional-motifs-in-intrinsically-unfolded-

INTECH
open science | open minds