Biochimica et Biophysica Acta 1830 (2013) 5236–5247

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

The conformational ensemble of the disordered and aggregation-protective 182–291 region of ataxin-3

Gaetano Invernizzi *[,1], Matteo Lambrughi [1], Maria Elena Regonesi, Paolo Tortora, Elena Papaleo *

Department of Biotechnology and Biosciences, University of Milano-Bicocca, 20126 Milan, Italy

ARTICLE INFO

Article history:
Received 5 March 2013
Received in revised form 10 June 2013
Accepted 10 July 2013
Available online 25 July 2013

Keywords:
Ataxin-3
Molecular dynamics
Polyglutamine
Aggregation
Ubiquitin interacting motif
Intrinsically disordered

ABSTRACT

Background: Intrinsically disordered proteins (IDPs) are an emerging part of structural biology that has challenged the classic paradigm of structure–function relationship. Indeed, IDPs have been associated with different physiological functions and associated with several pathologies, such as polyglutamine (polyQ) related diseases. Ataxin-3 (AT3) is the smallest polyQ protein, composed by the N-terminal folded Josephin domain (JD), which is amyloidogenic on its own, and a C-terminal unstructured part. The disordered region between the polyQ and the JD, $AT3_{182–291}$ plays a key role in the development of the disease.
Methods: We integrated different biophysical experimental techniques, atomistic explicit-solvent molecular dynamics (MD) simulations and graph theory to study $AT3_{182–291}$ structure.
Results: $AT3_{182–291}$ is a monomeric intrinsically disordered (ID) domain in solution and it is characterized by different conformational states, ascribable to pre-molten globule populations with different degrees of compactness. If isolated, it decreases the aggregation of the entire AT3.
Conclusions: We provided the first structural description of an ID domain associated to a polyQ protein and we also showed that it exerts protective effects against AT3 aggregation. This effect is likely to be induced by intermolecular interactions between AT3 and the ubiquitin-interacting motifs of $AT3_{182–291}$. Electrostatic interactions play a pivotal role in regulating the topology and tertiary propensity of the fragment and hub residues have been identified.
General significance: Synergistic use of atomistic simulations and biophysical techniques should be more generally applied to the study of IDPs. Since ID domains and polyQ-proteins are intimately connected, the study here provided can be of interest for other members of the group.

1. Introduction

Intrinsically disordered proteins (IDPs) have arisen in modern biology as one of the most interesting and unusual subjects to investigate. They constitute a large part of the polypeptides encoded by the genome and play important physiological roles exerted multi-functionally thanks to their (un)structural heterogeneity [1–3]. Further, more than one third of the human proteins contain sequences longer than 30 amino acids that are predicted to be disordered [4]. Indeed, several proteins have both folded and disordered domains. Many IDPs or proteins containing disordered domains are involved in a number of diseases, such as cancer, neurodegenerative and amyloidosis diseases [5,6]. Representative of this class of multi-domain proteins are polyglutamine proteins, a unique class of polypeptides with different functions and sequences that share the distinctive feature of a tract of consecutive glutamines located in disordered regions of the protein [7]. These multi-domain proteins have been extensively investigated since they trigger amyloid-related neurodegeneration upon expansion of the polyQ tract [8]. Notably, it has been demonstrated that sequences flanking the polyQ can deeply influence the aggregation pathway and consequently the toxicity of the protein [9–12]. Furthermore, it has been shown that early steps of the aggregation process are polyQ independent for ataxin-3 (AT3) and Huntingtin [13,14]. AT3 is one of the best characterized polyQ proteins, composed by a folded N-terminal domain and a disordered C-terminal part. The latter harbors the polyQ and two ubiquitin (Ub) interaction motifs (UIMs) necessary for the AT3 deubiquitinase activity [15–17]. The AT3 N-terminal folded domain, the so-called Josephin domain (JD) is the only AT3 domain for which structural information is available in atomic details [18,19]. JD is amyloidogenic on its own and it is responsible for the initial steps of AT3 aggregation [13]. Recently, we have demonstrated that the disordered part spanning from JD to polyQ (tract 182–291) increases the aggregation rate of JD and gives rise to aggregates with structural properties different from those of the isolated JD [20]. Furthermore, it has been demonstrated that an N-terminal truncated AT3 variant at residue 259, which is physiologically relevant as a cleavage product of

* Corresponding authors at. Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark. Tel.: +45 35322079
E-mail addresses: gaetano.invernizzi@unimib.it (G. Invernizzi), elena.papaleo@unimib.it, elena.papaleo@bio.ku.dk (E. Papaleo).
[1] The authors equally contributed to this work.

0304-4165/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbagen.2013.07.007

caspases, is toxic and induces an ataxic phenotype in a murine model [21].

In light of the above observations, the disordered tract between the JD and the polyQ seems to play a key role in the aggregation process and related toxicity of AT3. Interestingly, the disordered tract does not harbor any predicted aggregation prone region [20] nor does its presence alter the structural stability of the JD [22,23], leaving the open question of how it could influence JD aggregation. To shed light on this fundamental issue, an atomistic description of the conformational ensemble of the whole N-terminal part of AT3 is required to address the influence of the disordered tract on JD structural properties. The structural characterization of the disordered part thus becomes a mandatory step towards this goal.

Herein, we provide the first model of the conformational ensemble of the 183–291 region of AT3 in solution by atomistic explicit solvent multi-replicate simulations integrated to experimental data achieved by Circular Dichroism (CD), Size-Exclusion Chromatography (SEC) and ElectroSpray Ionization Mass Spectrometry (ESI-MS). It turns out to be an intrinsically disordered domain, characterized by α-helix structures in correspondence with the UIMs, as well as in other surrounding regions. The domain is likely to populate both poorly compact conformations with few long range intramolecular interactions and globular states mainly maintained by electrostatic interactions, most of which involve UIM elements.

2. Materials and methods

2.1. Analysis of the primary sequence

The net charge per residue was calculated according to Mao et al. [24]. Sequence hydrophobicity was calculated by the Kyte and Doolittle approximation by a window size of 5 amino acids and normalized in a scale between 0 and 1. The mean net charge per residue and mean hydrophobicity values were used to discriminate among intrinsically disordered and natively folded proteins, as proposed by Uversky et al. [25].

2.2. Molecular dynamics simulations

MD simulations were performed by Gromacs 4 (www.gromacs.org), implemented on a parallel architecture, with the GROMOS96 43a1. The starting models of the $AT3_{182–291}$ for MD simulations were generated by ab-initio modeling by I-Tasser (http://zhanglab.ccmb.med.umich.edu/I-TASSER/ [26]), also including the NMR structure of the AT3 UIMs in the Ub-bound conformation (PDB ID: 2KLZ) as one of the references. 35 models were generated, including distance restraints between residues L233-I253, D241-R250, I240-L249 (distances higher than 2 nm) to guide I-TASSER modeling and to avoid artificial intramolecular interactions in the models. We then selected as starting structure for the simulations one of the models that lacks sidechain–sidechain long-range intramolecular interactions. The model was soaked in a triclinic box of 19622 Simple Point Charge (SPC) water molecules with periodic boundary conditions. The box was built so that all the protein atoms were at a distance of at least 1.5 nm from the box edges. Starting from this system of 60123 atoms (protein plus water atoms) we carried out a first preparatory 50 ns MD simulation that was not included in the final ensemble. The final structure after 50 ns was used as a starting conformation for the subsequent MD simulations. In these MD simulations, the box was resized so that all the protein atoms were at a distance of at least 0.8 nm from the box edges to speed up the simulations (i.e. the total system, including both protein and water atoms was of 23586 atoms). In fact, after the first preparatory run the gyration radius of the protein decreases, as discussed in the Results and discussion section. Thus, we could afford to carry out the simulations using a lower number of water molecules. Productive MD simulations were performed in the isothermal–isobaric ensemble (300 K, 1 bar). The LINCS algorithm was employed to constrain heavy atom bond lengths, allowing for the use of a 2 fs time-step. Long-range electrostatic interactions were calculated using Particle-Mesh Ewald (PME) summation scheme. Van der Waals and Coulomb interactions were truncated at 1.2 nm, accordingly to cut-offs previously applied for simulations of IDPs and experimentally validated by comparison with electronic paramagnetic resonance and fluorescence data [27,28]. Na^+ and Cl^- counterions were added to the system to neutralize the overall charge and to simulate a physiological ionic strength (150 mM), according to a protocol previously employed for IDPs [28]. The non-pair list was updated every 10 steps and conformations were stored every 4 ps.

8 unconstrained MD simulations were carried out over 50 ns, extended to 100 ns in ad hoc selected cases to verify trajectories convergence, achieving overall more than 0.50 μs of MD ensemble overall. The time evolution of the main chain root mean square (rmsd) deviation with respect to the starting structure for the 8 MD runs was used to assess stability of each trajectory. The first 10 ns of each replicate, that features a major drift in main chain rmsd, were not included in the further analyses (*data not shown*). After evaluation of the achieved conformational sampling by principal components analysis (see below), the remaining 40 ns of each replicate was joined in a macro-trajectory for further analyses. In fact, the first four principal components of the macro-trajectory accounts for more than 70% of the *essential subspace*.

2.3. Analyses of the simulations

The solvent accessible surface (SAS) was calculated by g_sas Gromacs tool. The secondary structure (ss) content was calculated by DSSP program and g_helix Gromacs tool, along with a residue-dependent persistence degree of ss profile (pdssp).

Salt-bridges and their networks, along with networks of aromatic, amino-aromatic and hydrophobic interactions, were analyzed employing a persistence cut-off of 20% and a distance cutoff of 0.5 nm. Aromatic interactions were also verified at a higher cutoff of 0.6 nm. Hydrogen bonds were analyzed using a persistence cut-off of 20%, a distance cutoff between donor and acceptor group of 0.3 nm and a minimum donor–H-acceptor angle of 120°.

Salt-bridge networks were analyzed by the Pymol plugin xPyder [29]. xPyder represents pairwise relationships related to protein structures as two-dimensional matrices. In particular, the module for network analysis implemented in xPyder was employed. A network is described as a set of points (nodes) and connections between them (edges). A path is defined as a sequence of nodes for which an edge always exists between two consecutive nodes of the path. A matrix describing the persistence of each class of interactions was used as an input file. The program represents each residue of the matrix as a node of a simple, weighted graph connected by edges, whose weights are defined by the persistence of the interaction in the MD ensemble (i.e. the number of trajectory frames in which the interaction was present, over the total number of frames).

Residues connected by more than 3 edges to their neighbors are referred as hubs of the interaction network. The connected components, also called sub-networks, of the graph were also calculated. These are isolated sub-graphs in which all the edges are linked by at least one path, but no path exists between the nodes of the connected component and the rest of the graph. This analysis allows us to identify different clusters of interaction networks. The searching procedure was carried out so that the same node is not visited more than once to avoid entrapment in cycles.

2.4. Principal component analysis (PCA) and free energy landscape (FEL)

PCA highlights high-amplitude, concerted motions in MD trajectories, through the diagonalization of the mass-weighted covariance matrix (C) of the atomic positional fluctuations [30]. In particular, PCA analysis of MD trajectories provides a set of eigenvectors, each defined by an eigenvalue, describing the direction and the amplitude of the