

SHORT REPORT

Open Access

# Pathogenic polyglutamine expansion length correlates with polarity of the flanking sequences

Meewhi Kim<sup>1,2</sup>

## Abstract

**Background:** Polyglutamine (polyQ) repeat expansion within coding sequence of a soluble protein is responsible for eight autosomal-dominant genetic neurodegenerative disorders. These disorders affect cerebellum, striatum, basal ganglia and other brain regions. The pathogenic polyQ-expansion threshold in these proteins varies from 32Q to 54Q. Understanding the reasons for variability in pathogenic polyQ threshold may provide insights into pathogenic mechanisms responsible for development of these disorders.

**Findings:** Here we established a quantitative correlation between the polarity of the flanking sequences and pathogenic polyQ-expansion threshold in this protein family. We introduced an “edge polarity index” (<sup>E</sup>PI) to quantify polarity effects of the flanking regions and established a strong correlation between <sup>E</sup>PI index and critical polyQ expansion length in this protein family. Based on this analysis we subdivided polyQ-expanded proteins into 2 groups – with strong and weak dependence of polyQ threshold on <sup>E</sup>PI index. The main difference between members of the first and the second group is a polarity profile of these proteins outside of polyQ and flanking regions. PolyQ proteins are known substrates for proteasome and most likely mechanistic explanation for the observed correlation is that proteasome may have an impaired ability to process continuous non-polar regions of proteins.

**Conclusions:** The proposed hypothesis provides a quantitative explanation for variability in pathogenic threshold among polyQ-expansion disorders, which we established to correlate with polarity of flanking regions. To explain these results we propose that proteasome is not efficient in processing continuous non-polar regions of proteins, resulting in release of undigested and partially digested fragments. If supported experimentally, our hypothesis may have wide implications for further understanding the pathogenesis of polyglutamine expansion disorders.

**Keywords:** Polyglutamine disorders, Primary sequence analysis, Proteasome, Huntingtin, Ataxin

## Findings

### Introduction

An expansion of CAG-repeat encoding continuous stretch of glutamines results in polyQ expansion in the context of several proteins. These polyQ expansions are responsible for nine genetic neurodegenerative disorders affecting cerebellum, striatum, basal ganglia and other brain regions [1,2]. These disorders include Huntington's disease (HD), Dentatorubral-pallidolusian atrophy (DRPLA), Spinal Bulbar Muscular atrophy (SBMA), and spinocerebellar ataxias (SCAs) 1, 2, 3, 6, 7 and 17 [3-8] (Table 1). In the case of SCA6, the polyQ expansion

occurs in the context of a membrane protein – a pore forming subunit of Ca<sub>v</sub>2.1 voltage-gated calcium channel. For other 8 disorders the polyQ expansion occurs in the context of soluble proteins localized to the cytosol or the nucleus of neuronal cells. Each of these disorders has a characteristic pathogenic threshold for minimal polyQ expansion that causes disease [3-8]. Once this pathogenic threshold is exceeded, the age of disease onset is inversely proportional to polyQ expansion length [1]. With exception of the polyQ-expanded region, these proteins do not share other common sequence motifs or similar biological functions. All of these disorders are transmitted in an autosomal dominant fashion and the most likely reason for pathology is accumulation of polyQ-expanded proteins or protein fragments in cells. PolyQ-expanded proteins are able to interfere with multiple neuronal signaling pathways, such as protein aggregation, apoptotic cascade, calcium

Correspondence: MeeWhi.Kim@UTSouthwestern.edu

<sup>1</sup>Department of Physiology, UT Southwestern Medical Center, Dallas, TX 75390, USA

<sup>2</sup>Laboratory of Molecular Neurodegeneration, St Petersburg State Polytechnical University, St Petersburg 195251, Russian Federation

**Table 1 PolyQ-expansion diseases [1,2,17]**

Disease	Protein	Localization	$Q_{th}$	$^NPI$	$^CPI$	$^EPI$	Neuropathology
HD	Htt	Cytoplasm	35	7.925	-2.047	5.878	Striatum, Cortex
SBMA	AR	Nucleus, Cytoplasm	38	-3.658	19.475	15.817	Spinal cord, Brain stem
DRPLA	ATN1	Nucleus, Cytoplasm	49	53.530	36.706	90.236	Cerebellum, Basal ganglia
SCA1	Atxn1	Nucleus	39	35.109	25.323	60.432	Cerebellum, Brain stem
SCA2	Atxn2	Cytoplasm	32	13.614	-2.222	11.392	Cerebellum, Brain stem
SCA3	Atxn3	Nucleus, Cytoplasm	54	20.552	12.979	33.531	Ventral pons, Substantia nigra
SCA6	CACNA1A	Plasma Membrane	20	-	-	-	Cerebellum
SCA7	Atxn7	Nucleus, Cytoplasm	37	19.309	-2.047	17.262	Cerebellum, Brain stem, Spinal cord
SCA17	TBP	Nucleus	47	26.816	-3.758	23.058	Cerebellum, Striatum

Footnote:  $Q_{th}$  - pathogenic polyQ-expansion threshold;  $^NPI$  - amino terminal polarity index;  $^CPI$  - carboxy-terminal polarity index;  $^EPI$  - total edge polarity index.

signaling, proteasomal dysfunction, gene transcription and many others [9-16]. The interference with one or several of these signaling cascades eventually results in neuronal dysfunction and death, causing the disease. Pathogenic polyQ-expansion threshold ( $Q_{th}$ ) of soluble proteins in this gene family varies from 32Q (for SCA2) to 54Q (for SCA3) (Table 1). The reason for variability of the pathogenic threshold among these proteins is not understood. To explain this phenomenon, we performed comparative sequence analysis of the eight soluble polyQ-expanded proteins. Based on results of this quantitative analysis we established a correlation between the polarity of the flanking sequences and pathogenic polyQ-expansion threshold. To explain these results we propose that proteasome is not efficient in processing continuous non-polar regions of proteins, resulting in release of undigested and partially digested polyQ fragments.

## Methods

### Calculation of protein polarity profiles and polarity edge indexes

The Zimmerman Polarity index [18] for each protein was calculated by using ProtScale software package [19] with the following options: window size: 9, Relative weight of the window edges compared to the window center: 100%, Weight variation model: linear. The polarity edge index was derived by calculating change in polarity on the edges of polyQ sequence induced by the flanking regions. The amino terminal polarity index ( $^NPI$ ) and carboxy-terminal polarity index ( $^CPI$ ) are shown in Table 1. The total edge polarity index ( $^EPI$ ) was calculated as a sum of  $^NPI$  and  $^CPI$  indexes.

## Results

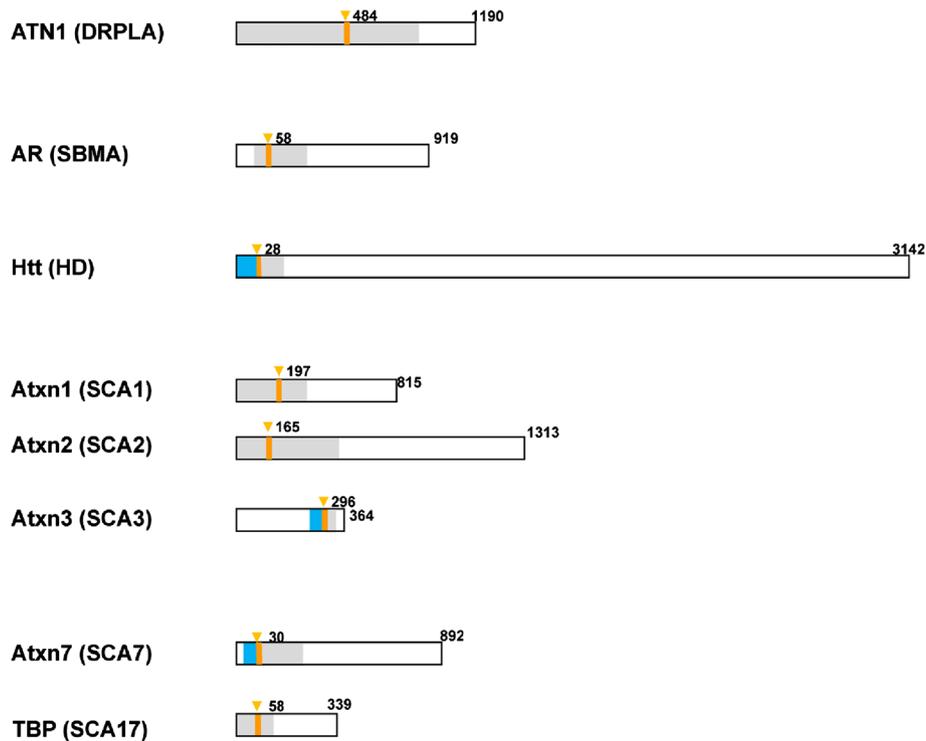
### Primary protein-sequence comparison of polyQ-expanded protein family

To determine potential sources of variability of pathogenic polyQ expansion threshold, we compared protein sequences of 8 soluble proteins in polyQ-expanded gene

family (Figure 1). The CACNA1 protein responsible for SCA6 was excluded from our analysis because CACNA1 is the only membrane protein in the family and because the polyQ-expansion threshold for SCA6 ( $Q_{th} = 20$ ) is much lower than those for all other polyQ-expansion disorders ( $Q_{th} > 35$ ). Most likely the pathogenic mechanism of SCA6 is distinct from all other disorders and directly related to changes in function or expression of  $Ca_v2.1$  voltage-gated  $Ca^{2+}$  channels [6]. The soluble members of polyQ-expanded protein family differ widely in size – from TBP (339 aa long) to Htt (3142 aa long) (Figure 1). The polyQ region (shown by orange bar) is located in the amino terminal portion of a protein (AR, Htt, Atxn1, Atxn2, Atxn7, TBP), in the central region of a protein (ATN1) or in the carboxy-terminal portion of a protein (Atxn3). Secondary structure predictions suggest that polyQ stretches in these proteins are flanked by a loop (grey) or  $\alpha$ -helical (blue) secondary structure. The presence of  $\alpha$ -helical flanking amino-terminal regions in Htt and Atxn3 proteins have been confirmed by crystallographic analysis ([20,21] and Meewhi Kim, unpublished observations). Crystallographic information for other polyQ-expanded proteins is not presently available. This initial analysis has not revealed any specific protein sequence pattern that could explain the variation in pathogenic polyQ-expansion threshold among the protein family (Table 1).

### Sequence polarity and proteasomal processing of protein

In cells, fragments of polyQ-expanded proteins are often found within ubiquitin-positive inclusions. These findings suggest that polyQ expansion may influence degradation of these proteins by proteasome [22-25]. Thus, we reasoned that the variability of pathogenic polyQ threshold may be due to variable effects of polyQ-flanking regions on proteasomal processing of these proteins. To better understand potential effects of protein sequence on proteasomal processing, we analyzed primary sequences



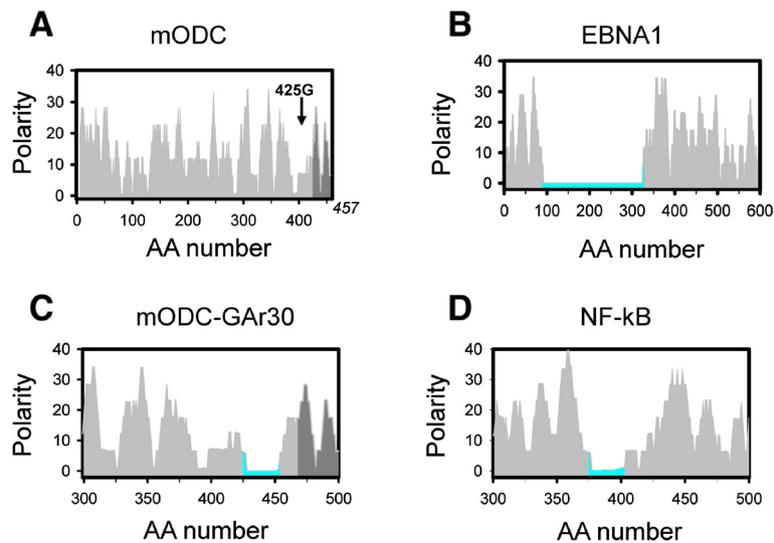
**Figure 1 The family of soluble polyQ-expanded proteins.** Schematic diagram of 8 soluble polyQ-expanded proteins. The last amino acid in each protein is numbered. The polyQ region is shown by an orange bar. The first residue of polyQ stretch is indicated by down triangle and numbered. The areas shaded by grey correspond to the loop regions based on secondary structure predictions. Predicted alpha-helical regions are shown by blue color.

of several model proteins which have been used previously as substrates for studies of proteasomal function. These proteins include mouse ornithine decarboxylase (mODC), which is degraded rapidly by the proteasome [26]. These proteins also included Epstein-Barr virus nuclear antigen1 (ENBA1) and Nuclear Factor kappa-light-chain-enhancer of activated B cells (NF-kB), both of which are extremely poor substrates for proteasomal degradation [27,28]. Proteasomal processing of these proteins is determined by their amino acid sequences. It has been previously suggested that the resistance to proteasomal processing is related to low sequence complexity of ENBA1 and NF-kB proteins [27,28]. We reasoned that sequence polarity may also affect proteasomal processing of these proteins. To evaluate this idea, we performed analysis of polarity profiles of these model proteasomal substrates. For mODC protein the polarity profile oscillates between 0 and 40 with periodicity between 5 to 20 amino acid residues through entire protein (Figure 2A). Such polarity profile appears to be optimal for proteasomal processing. In contrast, ENBA1 protein contains 250 amino acid long Glycine-Alanine repeats (GAR) with polarity close to 0 (Figure 2B). Similarly, NF-kB protein contains 30 amino acid long Glycine rich repeat region (GRR) of low polarity (Figure 2D). This analysis suggests that the presence of

extended low polarity region is incompatible with proteasomal processing of a substrate. Consistent with this interpretation, the insertion of a 30 amino acid long GAR repeat within mODC protein (ODC-GAR30) (Figure 2C) halted the proteasomal processing of the fusion protein, resulting in accumulation of undigested fragments of mODC-GAR30 in cells [29]. Based on this analysis we proposed that the presence of intermittent high/low polarity regions facilitates proteasomal degradation of proteins and the presence of extended low polarity regions impairs it due to premature release of incompletely digested substrate. This hypothesis is consistent with studies of dihydrofolate reductase (DHFR) degradation by proteasome [30]. Consistent with our hypothesis, insertion of low polarity stretches of 37S, 37N, or 37G aminoacids impaired ability of proteasome to process DHFR substrate in these experiments and lead to accumulation of partially digested products [30].

#### ***Polarity of polyQ-flanking regions and pathogenic polyQ threshold***

The expanded polyQ stretch corresponds to an extended low polarity sequence. As discussed above, we propose that such a sequence is a poor substrate for proteasomal processing. This argument may explain the accumulation



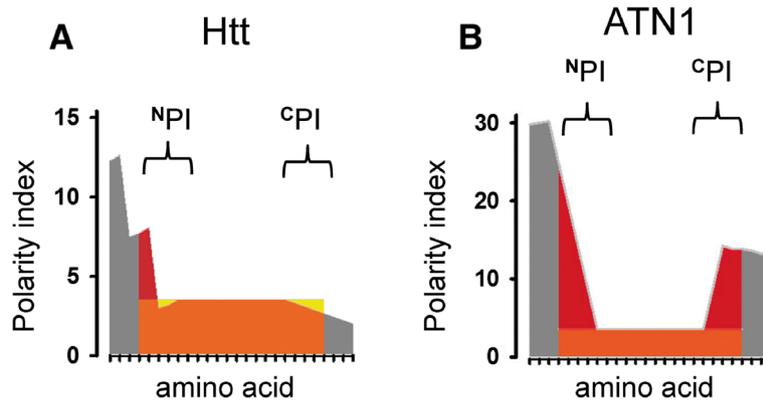
**Figure 2 Polarity profiles of model proteasomal substrates.** Calculated polarity profiles are shown for mODC (A); EBNA1 (B); ODC-GAr30 (C); NF-kB (D). Extended low polarity regions are colored with blue on panels B, C, D. In ODC-GAr30 construct the Gar30 repeat sequence was inserted in position 425G of mODC protein (shown on panel A).

of undigested ubiquitinated fragments of polyQ-expanded proteins in cells [23,24,31-36]. We further reasoned that the pathogenic threshold of polyQ expansion may be related to potential influence of flanking regions on proteasomal degradation of naked polyQ sequence. Specifically, if polyQ sequence is surrounded by highly polar flanking regions, then these regions can increase the effective polarity on the edges of polyQ stretch, promoting proteasomal processing. On the other hand, if polyQ region is flanked by low polarity regions, then the polarity of polyQ sequence edges remains low, impairing proteasomal processing. This hypothesis predicts that a polyQ sequence embedded within polar flanking sequences must have longer expansion to reach pathogenic threshold than the polyQ sequence embedded within low polarity flanking sequences.

To quantitatively test this hypothesis, we performed calculations of Edge Polarity Index ( $^EPI$ ) for each of the soluble polyQ-expanded proteins (Figure 1). The procedure used for these calculations is illustrated for Htt and ATN1 proteins on Figure 3. The polarity index of polyQ sequence (orange) is 3.5. Such value is indeed achieved in the middle of the polyQ sequence, away from the edges (Figures 3A,B). However, at the edges of polyQ sequence the polarity index deviates from the value of 3.5 due to the presence of the flanking regions (grey). If polarity of amino acids in the flanking regions exceeds 3.5, then the effective polarity on the edges of polyQ sequence is increased (shown by red color). However, if polarity of amino acids in the flanking regions is lower than 3.5, then effective polarity on the edges of polyQ sequence is reduced (shown by yellow color). To quantify these results, we calculated an integral change in polarity on the amino-terminal end ( $^NPI$ ) and

carboxy-terminal end ( $^CPI$ ) of the polyQ sequence. In these calculations a value of 3.5 was subtracted from the generated polarity index for each glutamine residue. The edge polarity index ( $^EPI$ ) was calculated as a sum of  $^NPI$  and  $^CPI$  values. Calculated  $^NPI$ ,  $^CPI$  and  $^EPI$  values are presented in Table 1 for each of the soluble polyQ-expanded proteins. The  $^EPI$  index for Htt is 5.878, one of the lowest in this protein family (Table 1). This is due to the presence of low polarity flanking sequence at the amino-termini of polyQ region and the presence of non-polar proline stretch at the carboxy-termini of the polyQ region (Figure 3A). In contrast, ATN1 has  $^EPI$  index of 90.236, the highest in this protein family (Table 1). This is due to the presence of highly polar flanking regions on both amino-terminal and carboxy-terminal ends of polyQ sequence (Figure 3B).

In order to test our hypothesis we aimed to establish the relationship between minimal pathogenic polyQ threshold and the polarity effects of the flanking sequences. When the length of polyQ pathogenic threshold was plotted versus  $^EPI$  for each polyQ-expanded protein, we observed a significant correlation (Figure 4, dashed line,  $r = 0.53$ ). The quality of the fit was improved significantly after separation of the eight soluble polyQ-expanded proteins into 2 groups. The first group includes Htt, Atxn7, AR, TBP and Atxn3 (Figure 4, filled squares,  $r = 0.94$ ). For this group polyQ pathogenic threshold increases as a strong function of the  $^EPI$  index (Figure 4). The second group includes Atxn2, Atxn1, and ATN1 (Figure 4, open squares,  $r = 0.97$ ). For these proteins polyQ pathogenic threshold increases as a weak function of the  $^EPI$  index (Figure 4).



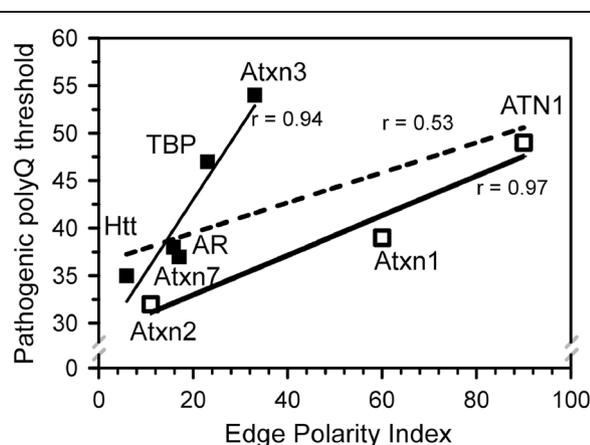
**Figure 3 Calculation of Edge Polarity Index.** The polarity profile is shown for the polyQ stretch and immediate flanking regions for Htt (A) and ATN1 (B). The polarity index of naked polyQ stretch (3.5) is shown by orange. The polarity index of flanking regions are shown by grey. Effective increase of polarity index on the edges of polyQ sequence is shown by red. Effective decrease of polarity index on the edges of polyQ sequence is shown by yellow. <sup>N</sup>PI and <sup>C</sup>PI indexes were calculated by integrating edge polarity effects at amino-terminal and carboxy-terminal ends of polyQ sequence.

#### Pathogenic polyQ threshold and influence of protein context

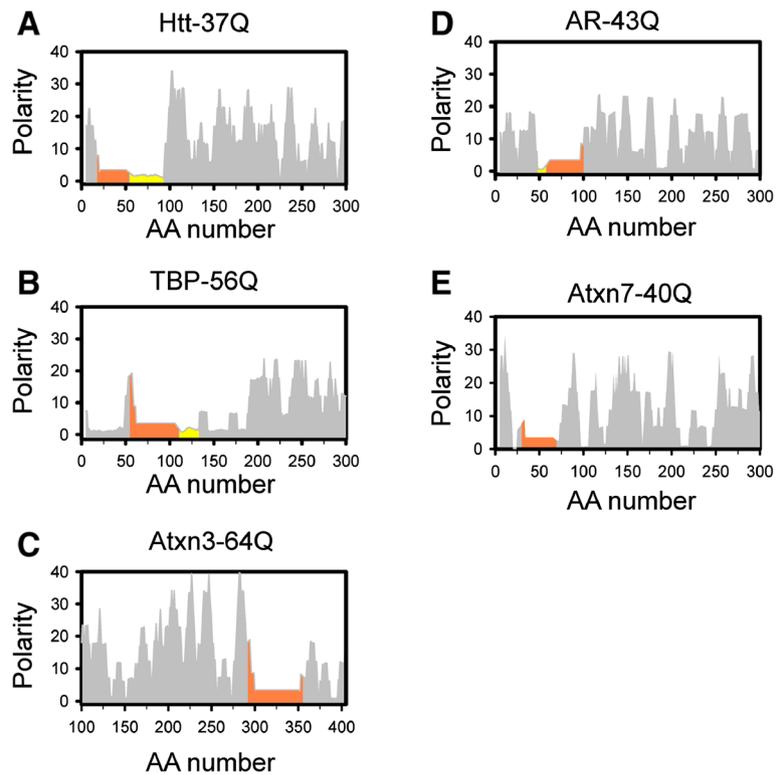
To understand the difference between the proteins in the first and second groups we analyzed the polarity profiles for each of these proteins in more detail. Whole protein sequence was analyzed for TBP protein, the shortest member in this family. For the remaining 7 proteins we focused on the 300 amino-acid region that included polyQ sequence. The main reason to focus on just 300 amino-acid region is that release of undigested polyQ-containing fragment has most important implications for toxicity of these proteins in cells when compared to remaining portions of these proteins. Calculated polarity profiles for

5 members of the first group are shown on Figure 5. Htt protein contains an extended low-polarity proline-rich flanking region on carboxy-terminal side of the polyQ stretch (Figure 5A, shown in yellow). Due to the presence of this low polarity region pathogenic polyQ threshold for HD ( $Q_{th} = 35$ ) is one of the shortest in this family of proteins (Table 1). With the exception of polyQ and flanking regions the rest of the Htt N-terminal region is composed of regular high-low polarity residues. The polyQ-flanking domains in TBP are a polar peak at N-terminal and a short non-polar proline-rich domain at C-terminal (Figure 5B). Atxn3 contains regular polar flanking regions at both sides (Figure 5C). Sequences of AR and Atxn7 are composed of residues with regular polar regions (Figures 5D,E). Therefore, with exception of the polyQ stretches and flanking regions, all members of this group are composed of regular low-high polarity sequences (Figures 5A-E). We suggest that such sequences are good substrates for proteasomal degradation. The polarity plots for 3 members of the second group (Atxn2, Atxn1, ATN1) are shown in Figure 6. These 3 proteins are composed primarily of low polarity regions (Figures 6A-C). We suggest that such sequences are poor substrates for proteasomal degradation. Based on this analysis we propose that the difference between the members of the first and the second groups come from the overall polarity profiles of these proteins beyond polyQ stretch and immediate flanking sequences.

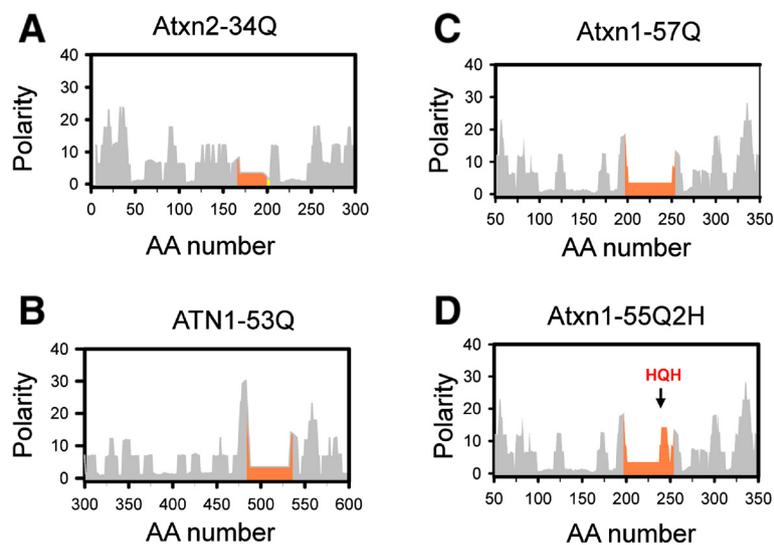
Interestingly, each member of the first group has been reported to be associated with proteasome in biochemical experiments. In some cases association with proteasome occurred via ubiquitinated form of the protein - such as AR [37,38], Atxn3 [39,40], or Htt [41,42]. In some cases association with proteasome did not require ubiquitination, such as for Atxn7 [43,44] and for TBP [45-47]. In contrast, members of the second group have not been



**Figure 4 Correlation between pathogenic polyQ expansion threshold and Edge Polarity Index.** Pathogenic polyQ-expansion threshold is plotted versus <sup>E</sup>PI index for each of the 8 soluble polyQ-expanded proteins. Linear fit for all 8 data points is shown by a dashed line ( $r = 0.53$ ). The linear fit for group 1 (filled squares,  $r = 0.94$ ) and group 2 (open squares,  $r = 0.97$ ) members is shown by solid lines.



**Figure 5 Polarity profiles of Group 1 polyQ-expanded proteins.** Polarity index profiles are shown for Htt-37Q (A); TBP-56Q (B); Atxn3-64Q (C); AR-43Q (D); Atxn7-40Q (E). PolyQ region in each protein is shown by orange. The proline-rich regions are shown by yellow. Full-length protein profile is shown for TBP protein. The profiles for 300 amino-acid long protein fragments are shown for other proteins.



**Figure 6 Polarity profiles of Group 2 polyQ-expanded proteins.** Polarity index profiles are shown for Atxn2-34Q (A); ATN1-53Q (B); Atxn1-57Q (C) and Atxn1-55Q2H (D). PolyQ region in each protein is shown by orange. The profiles for 300 amino-acid long protein fragments are shown for each protein. The site of HQH insertion is shown by arrow on panel D.

reported to associate with proteasome in biochemical experiments. The only known interaction is proteasomal association of Atxn1 that is mediated by HSP/CHIP [48,49] and requires partial unfolding of Atxn1 to be initiated. Although correlative, this argument further suggests that the members of the first group are better substrates for proteasomal degradation than the members of the second group.

#### **Effects of Histidine insertion**

A unique clinical case provides an indirect support to our hypothesis. A Japanese SCA1 patient was discovered to have an insertion of 2 His residues within polyQ stretch, resulting in sequence Q<sub>45</sub>HQH<sub>10</sub> [50]. An expected age of disease onset for a typical SCA1-58Q patient is 22 years of age. In contrast, the SCA1-Q<sub>45</sub>HQH<sub>10</sub> patient displayed first symptoms of disease at the age of 50 [50]. In addition, the brain stem atrophy of this patient was much milder than expected for a typical SCA1 patient with similar repeat length [50]. What is an explanation for dramatic protective effects of His insertion? Biophysical studies [51-53] and our own crystallographic experiments [21] suggested that insertion of His has minimal effect on secondary structure of polyQ region. Thus, it is not likely that insertion of 2 His residues disrupted the “toxic conformation” of the 58Q stretch. However, insertion of 2 His residues is expected to introduce a polarity peak within polyQ sequence. Indeed, the polarity profile of Atxn1-55Q2H (Q<sub>45</sub>HQH<sub>10</sub>) contains a significant polar peak (Figure 6D). We propose that such polar insertion enhances proteasomal processing of His-containing protein. As a result, “effective” low polarity polyQ region is shortened to approximately 36Q (Figure 6D), which is consistent with the very mild clinical phenotype of this particular patient [50].

#### **Conclusion**

In this paper we established a quantitative correlation between the polarity of the flanking regions and the pathogenic polyQ expansion threshold for the soluble polyQ-containing proteins. The quantitative analysis enabled us to divide soluble polyQ-expanded proteins into 2 groups – with strong and weak dependence of polyQ threshold on the polarity of the flanking regions. The main difference between members of the first group (Htt, Atxn7, AR, TBP, Atxn3) and the second group (Atxn2, Atxn1, ATN1) is related to polarity profile of remainder of these proteins. All members of the first group composed of regular low-high polarity sequences, whereas members of the second group are composed primarily of low polarity sequence regions. PolyQ proteins are known substrates for proteasomal degradation. We analyzed experiments performed with the model proteasomal substrates and concluded that proteasome

has impaired ability to process continuous non-polar regions of proteins. We propose that polarity of flanking regions may have an important modulatory effects on ability of proteasome to process continuous polyQ stretches, resulting in accumulation of polyQ-expanded proteins and partially digested protein fragments in cells. Such proteins can then exert “toxic gain of function” effects by interfering with essential neuronal signaling pathways. These ideas need to be tested experimentally. However, indirect support to our hypothesis is provided by partial protective effects of His insertion within polyQ stretch of SCA1 patient, which has a significant effect on polarity profile of polyQ stretch. We propose that such polar insertion can facilitate proteasomal degradation of polyQ-expanded Ataxin 1, which may explain less severe pathology in these patients.

#### **Competing interests**

The author declares that she has no competing interest.

#### **Acknowledgments**

I thank to Leah Taylor for administrative assistance, members of Bezprozvanny and Kim laboratories for helpful discussions and Drs Ilya Bezprozvanny (I.B.) and George De Martino for comments on the paper. MWK is a Young Investigator of the National Ataxia Foundation and supported by the Hereditary Disease Foundation. This work was also supported by the contract with the Russian Ministry of Science 11. G34.31.0056 (I.B.), by the Russian Scientific Fund grant 14-25-00024 (I.B.), and by the NIH grants R01NS074376 and R01NS056224 (I.B.).

Received: 13 June 2014 Accepted: 23 October 2014

Published: 6 November 2014

#### **References**

1. Gusella JF, MacDonald ME: **Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease.** *Nat Rev Neurosci* 2000, **1**:109–115.
2. Havel LS, Li S, Li XJ: **Nuclear accumulation of polyglutamine disease proteins and neuropathology.** *Mol Brain* 2009, **2**:21.
3. Banfi S, Chung MY, Kwiatkowski TJ Jr, Ranum LP, McCall AE, Chinault AC, Orr HT, Zoghbi HY: **Mapping and cloning of the critical region for the spinocerebellar ataxia type 1 gene (SCA1) in a yeast artificial chromosome contig spanning 1.2 Mb.** *Genomics* 1993, **18**:627–635.
4. Pulst SM, Nechiporuk A, Nechiporuk T, Gispert S, Chen XN, Lopes-Cendes I, Pearlman S, Starkman S, Orozco-Diaz G, Lunke A, DeJong P, Rouleau GA, Auburger G, Korenberg JR, Figueroa C, Sahba S: **Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2.** *Nat Genet* 1996, **14**:269–276.
5. Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, Nakamura S, Nishimura M, Akiguchi I, Kimura J, Narumiya S, Kakizuka A: **CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1.** *Nat Genet* 1994, **8**:221–228.
6. Toru S, Murakoshi T, Ishikawa K, Saegusa H, Fujigasaki H, Uchihara T, Nagayama S, Osanai M, Mizusawa H, Tanabe T: **Spinocerebellar ataxia type 6 mutation alters P-type calcium channel function.** *J Biol Chem* 2000, **275**:10893–10898.
7. David G, Abbas N, Stevanin G, Durr A, Yvert G, Cancel G, Weber C, Imbert G, Saudou F, Antoniou E, Drabkin H, Gemmill R, Giunti P, Benomar A, Wood N, Ruberg M, Agid Y, Mandel JL, Brice A: **Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion.** *Nat Genet* 1997, **17**:65–70.
8. Nakamura K, Jeong SY, Uchihara T, Anno M, Nagashima K, Nagashima T, Ikeda S, Tsuji S, Kanazawa I: **SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein.** *Hum Mol Genet* 2001, **10**:1441–1448.
9. Tobin AJ, Signer ER: **Huntington's disease: the challenge for cell biologists.** *Trends Cell Biol* 2000, **10**:531–536.

10. Ross CA: Polyglutamine pathogenesis: emergence of unifying mechanisms for Huntington's disease and related disorders. *Neuron* 2002, **35**:819–822.
11. Rubinsztein DC: Lessons from animal models of Huntington's disease. *Trends Genet* 2002, **18**:202–209.
12. Li SH, Li XJ: Huntingtin-protein interactions and the pathogenesis of Huntington's disease. *Trends Genet* 2004, **20**:146–154.
13. Bezprozvanny I: Calcium signaling and neurodegenerative diseases. *Trends Mol Med* 2009, **15**:89–100.
14. Cha JH: Transcriptional signatures in Huntington's disease. *Prog Neurobiol* 2007, **83**:228–248.
15. Truant R, Atwal RS, Desmond C, Munsie L, Tran T: Huntington's disease: revisiting the aggregation hypothesis in polyglutamine neurodegenerative diseases. *FEBS J* 2008, **275**:4252–4262.
16. Takahashi T, Katada S, Onodera O: Polyglutamine diseases: where does toxicity come from? what is toxicity? where are we going? *J Mol Cell Biol* 2010, **2**:180–191.
17. Pennuto M, Palazzolo I, Poletti A: Post-translational modifications of expanded polyglutamine proteins: impact on neurotoxicity. *Hum Mol Genet* 2009, **18**:R40–R47.
18. Zimmerman JM, Eliezer N, Simha R: The characterization of amino acid sequences in proteins by statistical methods. *J Theor Biol* 1968, **21**:170–201.
19. Gasteiger EHC, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A: Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook*. Edited by John M, Walker JM. New York City, USA: Humana Press; 2005:571–607.
20. Kim MW, Chelliah Y, Kim SW, Otwinowski Z, Bezprozvanny I: Secondary structure of Huntingtin amino-terminal region. *Structure* 2009, **17**:1205–1212.
21. Kim M: Beta conformation of polyglutamine track revealed by a crystal structure of Huntingtin N-terminal region with insertion of three histidine residues. *Prion* 2013, **7**:221–228.
22. Raspe M, Gillis J, Krol H, Krom S, Bosch K, van Veen H, Reits E: Mimicking proteasomal release of polyglutamine peptides initiates aggregation and toxicity. *J Cell Sci* 2009, **122**:3262–3271.
23. Holmberg CI, Staniszewski KE, Mensah KN, Matouschek A, Morimoto RI: Inefficient degradation of truncated polyglutamine proteins by the proteasome. *EMBO J* 2004, **23**:4307–4318.
24. Venkatraman P, Wetzel R, Tanaka M, Nukina N, Goldberg AL: Eukaryotic proteasomes cannot digest polyglutamine sequences and release them during degradation of polyglutamine-containing proteins. *Mol Cell* 2004, **14**:95–104.
25. Jana NR, Zemskov EA, Wang G, Nukina N: Altered proteasomal function due to the expression of polyglutamine-expanded truncated N-terminal huntingtin induces apoptosis by caspase activation through mitochondrial cytochrome c release. *Hum Mol Genet* 2001, **10**:1049–1059.
26. Ghoda L, van Daalen Wetters T, Macrae M, Ascherman D, Coffino P: Prevention of rapid intracellular degradation of ODC by a carboxyl-terminal truncation. *Science* 1989, **243**:1493–1495.
27. Blake NW, Moghaddam A, Rao P, Kaur A, Glickman R, Cho YG, Marchini A, Haigh T, Johnson RP, Rickinson AB, Wang F: Inhibition of antigen presentation by the glycine/alanine repeat domain is not conserved in simian homologues of Epstein-Barr virus nuclear antigen 1. *J Virol* 1999, **73**:7381–7389.
28. Lin L, Ghosh S: A glycine-rich region in NF-kappaB p105 functions as a processing signal for the generation of the p50 subunit. *Mol Cell Biol* 1996, **16**:2248–2254.
29. Hoyt MA, Zich J, Takeuchi J, Zhang M, Govaerts C, Coffino P: Glycine-alanine repeats impair proper substrate unfolding by the proteasome. *EMBO J* 2006, **25**:1720–1729.
30. Tian L, Holmgren RA, Matouschek A: A conserved processing mechanism regulates the activity of transcription factors Cubitus interruptus and NF-kappaB. *Nat Struct Mol Biol* 2005, **12**:1045–1053.
31. DiFiglia M, Sapp E, Chase KO, Davies SW, Bates GP, Vonsattel JP, Aronin N: Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* 1997, **277**:1990–1993.
32. Haacke A, Bradley SA, Boteva R, Tzvetkov N, Hartl FU, Breuer P: Proteolytic cleavage of polyglutamine-expanded ataxin-3 is critical for aggregation and sequestration of non-expanded ataxin-3. *Hum Mol Genet* 2006, **15**:555–568.
33. Landles C, Sathasivam K, Weiss A, Woodman B, Moffitt H, Finkbeiner S, Sun B, Gafni J, Ellerby LM, Trotter Y, Richards WG, Osmand A, Paganetti P, Bates GP: Proteolysis of mutant huntingtin produces an exon 1 fragment that accumulates as an aggregated protein in neuronal nuclei in Huntington disease. *J Biol Chem* 2010, **285**:8808–8823.
34. Young JE, Gouw L, Propp S, Sopher BL, Taylor J, Lin A, Hermel E, Logvinova A, Chen SF, Chen S, Bredesen DE, Truant R, Ptacek LJ, La Spada AR, Ellerby LM: Proteolytic cleavage of ataxin-7 by caspase-7 modulates cellular toxicity and transcriptional dysregulation. *J Biol Chem* 2007, **282**:30150–30160.
35. Suzuki Y, Nakayama K, Hashimoto N, Yazawa I: Proteolytic processing regulates pathological accumulation in dentatorubral-pallidoluysian atrophy. *FEBS J* 2010, **277**:4873–4887.
36. Verhoef LG, Lindsten K, Masucci MG, Dantuma NP: Aggregate formation inhibits proteasomal degradation of polyglutamine proteins. *Hum Mol Genet* 2002, **11**:2689–2700.
37. Beitel LK, Elhaji YA, Lumbroso R, Wing SS, Panet-Raymond V, Gottlieb B, Pinsky L, Trifiro MA: Cloning and characterization of an androgen receptor N-terminal-interacting protein with ubiquitin-protein ligase activity. *J Mol Endocrinol* 2002, **29**:41–60.
38. Poukka H, Karvonen U, Janne OA, Palvimo JJ: Covalent modification of the androgen receptor by small ubiquitin-like modifier 1 (SUMO-1). *Proc Natl Acad Sci U S A* 2000, **97**:14145–14150.
39. Burnett B, Li F, Pittman RN: The polyglutamine neurodegenerative protein ataxin-3 binds polyubiquitylated proteins and has ubiquitin protease activity. *Hum Mol Genet* 2003, **12**:3195–3205.
40. Kuhlbrodt K, Janiesch PC, Kevei E, Segref A, Barikbin R, Hoppe T: The Machado-Joseph disease deubiquitylase ATX-3 couples longevity and proteostasis. *Nat Cell Biol* 2011, **13**:273–281.
41. Kalchman MA, Graham RK, Xia G, Koide HB, Hodgson JG, Graham KC, Goldberg YP, Gietz RD, Pickart CM, Hayden MR: Huntingtin is ubiquitinated and interacts with a specific ubiquitin-conjugating enzyme. *J Biol Chem* 1996, **271**:19385–19394.
42. Steffan JS, Agrawal N, Pallos J, Rockabrand E, Trotman LC, Slepko N, Illes K, Lukacsovich T, Zhu YZ, Cattaneo E, Pandolfi PP, Thompson LM, Marsh JL: SUMO modification of Huntingtin and Huntington's disease pathology. *Science* 2004, **304**:100–104.
43. Ansong O, Giunti P, Michalik A, Van Broeckhoven C, Harding B, Wood N, Scaravilli F: Ataxin-7 aggregation and ubiquitination in infantile SCA7 with 180 CAG repeats. *Ann Neurol* 2004, **56**:448–452.
44. Matilla A, Gorbea C, Einum DD, Townsend J, Michalik A, van Broeckhoven C, Jensen CC, Murphy KJ, Ptacek LJ, Fu YH: Association of ataxin-7 with the proteasome subunit S4 of the 19S regulatory complex. *Hum Mol Genet* 2001, **10**:2821–2831.
45. Ferrell K, Wilkinson CR, Dubiel W, Gordon C: Regulatory subunit interactions of the 26S proteasome, a complex problem. *Trends Biochem Sci* 2000, **25**:83–88.
46. Chew BS, Siew WL, Xiao B, Lehming N: Transcriptional activation requires protection of the TATA-binding protein Tbp1 by the ubiquitin-specific protease Ubp3. *Biochem J* 2010, **431**:391–399.
47. Trachtulec Z, Hamvas RM, Forejt J, Lehrach HR, Vincek V, Klein J: Linkage of TATA-binding protein and proteasome subunit C5 genes in mice and humans reveals synteny conserved between mammals and invertebrates. *Genomics* 1997, **44**:1–7.
48. Al-Ramahi I, Lam YC, Chen HK, de Gouyon B, Zhang M, Perez AM, Branco J, de Haro M, Patterson C, Zoghbi HY, Botas J: CHIP protects from the neurotoxicity of expanded and wild-type ataxin-1 and promotes their ubiquitination and degradation. *J Biol Chem* 2006, **281**:26714–26724.
49. Choi JY, Ryu JH, Kim HS, Park SG, Bae KH, Kang S, Myung PK, Cho S, Park BC, do Lee H: Co-chaperone CHIP promotes aggregation of ataxin-1. *Mol Cell Neurosci* 2007, **34**:69–79.
50. Matsuyama Z, Izumi Y, Kameyama M, Kawakami H, Nakamura S: The effect of CAT trinucleotide interruptions on the age at onset of spinocerebellar ataxia type 1 (SCA1). *J Med Genet* 1999, **36**:546–548.
51. Jayaraman M, Kodali R, Wetzel R: The impact of ataxin-1-like histidine insertions on polyglutamine aggregation. *Protein Eng Des Sel* 2009, **22**:469–478.
52. Sharma D, Sharma S, Pasha S, Brahmachari SK: Peptide models for inherited neurodegenerative disorders: conformation and aggregation properties of long polyglutamine peptides with and without interruptions. *FEBS Lett* 1999, **456**:181–185.
53. Sen S, Dash D, Pasha S, Brahmachari SK: Role of histidine interruption in mitigating the pathological effects of long polyglutamine stretches in SCA1: a molecular approach. *Protein Sci* 2003, **12**:953–962.

doi:10.1186/1750-1326-9-45

Cite this article as: Kim: Pathogenic polyglutamine expansion length correlates with polarity of the flanking sequences. *Molecular Neurodegeneration* 2014 **9**:45.