

# Polyglutamine Induced Misfolding of Huntingtin Exon1 is Modulated by the Flanking Sequences

Vinal V. Lakhani, Feng Ding, Nikolay V. Dokholyan\*

Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, School of Medicine, Chapel Hill, North Carolina, United States of America

## Abstract

Polyglutamine (polyQ) expansion in exon1 (XN1) of the huntingtin protein is linked to Huntington's disease. When the number of glutamines exceeds a threshold of approximately 36–40 repeats, XN1 can readily form amyloid aggregates similar to those associated with disease. Many experiments suggest that misfolding of monomeric XN1 plays an important role in the length-dependent aggregation. Elucidating the misfolding of a XN1 monomer can help determine the molecular mechanism of XN1 aggregation and potentially help develop strategies to inhibit XN1 aggregation. The flanking sequences surrounding the polyQ region can play a critical role in determining the structural rearrangement and aggregation mechanism of XN1. Few experiments have studied XN1 in its entirety, with all flanking regions. To obtain structural insights into the misfolding of XN1 toward amyloid aggregation, we perform molecular dynamics simulations on monomeric XN1 with full flanking regions, a variant missing the polyproline regions, which are hypothesized to prevent aggregation, and an isolated polyQ peptide ( $Q_n$ ). For each of these three constructs, we study glutamine repeat lengths of 23, 36, 40 and 47. We find that polyQ peptides have a positive correlation between their probability to form a  $\beta$ -rich misfolded state and their expansion length. We also find that the flanking regions of XN1 affect its probability to form a  $\beta$ -rich state compared to the isolated polyQ. Particularly, the polyproline regions form polyproline type II helices and decrease the probability of the polyQ region to form a  $\beta$ -rich state. Additionally, by lengthening polyQ, the first N-terminal 17 residues are more likely to adopt a  $\beta$ -sheet conformation rather than an  $\alpha$ -helix conformation. Therefore, our molecular dynamics study provides a structural insight of XN1 misfolding and elucidates the possible role of the flanking sequences in XN1 aggregation.

**Citation:** Lakhani VV, Ding F, Dokholyan NV (2010) Polyglutamine Induced Misfolding of Huntingtin Exon1 is Modulated by the Flanking Sequences. PLoS Comput Biol 6(4): e1000772. doi:10.1371/journal.pcbi.1000772

**Editor:** Ruth Nussinov, National Cancer Institute United States of America and Tel Aviv University, Israel

**Received:** November 22, 2009; **Accepted:** March 30, 2010; **Published:** April 29, 2010

**Copyright:** © 2010 Lakhani et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work is supported by the National Institute of Health Grant R01GM080742 and the ARRA supplement 3R01GM080742-03S1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: dokh@med.unc.edu

## Introduction

Similar to eight other neurodegenerative polyglutamine diseases [1–5], Huntington's disease (HD) is associated with a polyglutamine (polyQ) expansion in the huntingtin protein. In HD, the first exon (XN1) of the huntingtin protein, which contains the polyQ region, has been implicated as the pathogenic polypeptide [6–9]. It is hypothesized that after it is cleaved from huntingtin, the XN1 polypeptides misfold and form lethal cellular aggregates or inclusions in neuronal cells [1–5]. Evidence from *in vivo* studies [6,7] indicates that XN1 polypeptides can form aggregates similar to those observed in the neurons of afflicted patients. Despite limited knowledge on the structure and function of XN1, some information about its aggregation is known. The length of the polyQ region in XN1 is inversely proportional to the age of onset of symptoms [2,9]. Additionally, there is a threshold of approximately 36–40 repeats, within which symptoms may or may not develop [10]. However, beyond this threshold, lethal symptoms eventually develop in the lifetime of the patient. These symptoms develop earlier in the life of patients with longer repeat lengths. *In vitro* kinetic studies of polyQ peptides have emulated this clinical correlation between age and length [11,12]. Kinetic studies of polyQ [13] aggregation suggested that the misfolding of a polyQ monomer initiates the aggregation. Although other studies [14,15] proposed more complex aggregation scenario, we hypothesize that

the misfolding of a polyQ monomer plays a critical role in the formation of ordered amyloid aggregations. Similarly, the aggregation of XN1 displays more complicated aggregation behavior [16–18,19,20], it has been shown that the length-dependent misfolding of polyQ sequence in the XN1 monomer plays a critical role in the initial oligomerization and the later formation of  $\beta$ -rich amyloid aggregates [20]. Here, we propose to study the structure and dynamics of monomeric XN1, in hopes to help illuminate the aggregation mechanism and help determine the role of these aggregates.

Due to the structural complexity of XN1, previous experimental and computational studies have focused on specific regions of XN1 (Fig. 1). First, by studying polyQ homopolymers ( $Q_n$ ), and similar constructs, it has been shown that as the length increases,  $Q_n$  transitions from forming random coils in solution to  $\beta$ -sheets [16,21–27]. Yet, at least one computational study finds no such transition [28]. Other authors find the formation of a  $\beta$ -helix [29–33], first proposed by Perutz [34]. Additionally, the polyQ region has been shown to unfold neighboring regions [20,35–38], which suggests the flanking regions should be included to better model the behavior of XN1. Second, others [25,39–41] have studied polyQ together with a polyproline domain, because XN1 has two polyproline (polyP) regions near the polyQ region (Fig. 1a). Experimental studies of these models, indicate that the polyP regions form polyproline type II (PPII) helices [3,40,41], which

## Author Summary

Huntington's Disease is a neurodegenerative disorder associated with protein aggregation in neurons. The aggregates formed are thought to lead to neurotoxicity and cell death. Understanding the molecular structure of these aggregates may lead to strategies to inhibit aggregation. Exon 1 (XN1) of the huntingtin protein is critical for aggregate formation. This polypeptide has a naturally occurring polyglutamine sequence (polyQ), which is elongated in patients afflicted with the disease. The polyQ region in XN1 has several flanking sequences with distinct physicochemical properties, including the N-terminal 17 residues, two polyproline regions, and C-terminal sequences, that may affect its overall structure and aggregation. What is the overall structure of XN1, and what structural effects do the neighboring sequences have on each other and polyQ? We address these questions by studying computational models of various polypeptides, including XN1 and three mutant forms associated with Huntington's Disease. Certain neighboring sequences are found to inhibit aggregation, while others may be recruited by polyQ to form aggregates. Our results suggest the role that the flanking sequences may play in XN1 aggregation and may subsequently guide future structural models of XN1 aggregation.

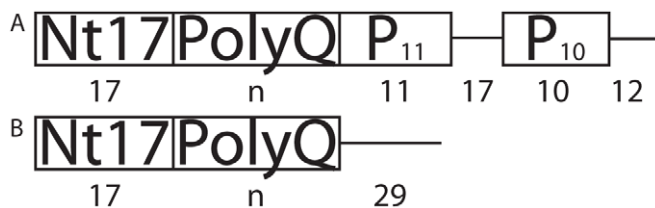
compete with the  $\beta$ -sheet structure of the polyQ region [40] and possibly protect against aggregation [1,3,5,20,39,40,42–44]. Last, recent models that include the first 17 residues (Nt17) of XN1 that are N terminal to the polyQ region have been studied [20,42,45–47]. By studying isolated Nt17 polypeptides, investigators have found that this region is either a compact random coil [20], an  $\alpha$ -helix [42,45] or both [20,46], which resists aggregation [20,46]. However, an Nt17+polyQ chimeric polypeptide was found to be highly prone to aggregation, more so than an unflanked  $Q_n$  homopolymer [20]. Thus, the structural role of the Nt17 region in XN1 aggregation is still unclear. Although the polyQ region has been individually shown to aggregate through studies on  $Q_n$ , its effects are unknown in the context of the XN1 flanking regions. For example, it is uncertain whether the Nt17 region assists [3,20,45] or hinders [1,46,48] aggregation, and perhaps separately, to what extent the polyP region prevents aggregation. Therefore, it is necessary to study XN1 misfolding and aggregation in the context of flanking regions.

Despite the structural information about each region, there still lacks a complete picture of XN1 due to the non-additive effect of

interactions between different structural elements. Here, we perform all-atom discrete molecular dynamics (DMD) [49–52] simulations to systematically study structural dynamics of XN1 and its variants. DMD has been shown to have a higher sampling efficiency than traditional molecular dynamics and has been used to study protein folding thermodynamics and protein aggregation [53]. All-atom DMD features a transferable force field and has been successfully used to fold several small proteins *ab initio* [52] and to study the folding and misfolding dynamics of Cu, Zn superoxide dismutase [54]. We construct 12 polypeptides (Fig. 1) categorized into three sets. In the first set, we study XN1 in its entirety (Fig. 1a) in order to capture the interactions between the Nt17, polyQ and polyP flanking regions simultaneously. In the second set, we study XN1 without the polyP flanking regions, titled XN1-P<sub>11</sub>-P<sub>10</sub>, (Fig. 1b) in order to determine the effect of the polyP regions on XN1 in the context of naturally occurring flanking regions. In the third set, we study polyQ homopolymers in the absence of all flanking regions ( $Q_n$ ) as controls. In each set, four different numbers of glutamine repeats are included: 23 (non-pathogenic), 36 (threshold), 40 and 47 (pathogenic). As originally suggested by Perutz [36] and expanded upon by others [1,3,20,35,36,38,55], the sequence context for a polyQ region can alter its aggregation mechanism; here, we find a more complete view of how the context plays a significant role in the secondary structure. In the context of XN1, the residues in the polyQ region have a lower probability of adopting  $\beta$ -sheet conformations, due to inhibition by the polyP regions. Surprisingly, by increasing the number of glutamine repeats in the polyQ region, the Nt17 region can be induced to fold into a  $\beta$ -strand. Thus, we suggest that the polyQ and flanking regions in XN1 are strongly coupled in XN1 folding and misfolding.

## Results

For each of the 12 models (Fig. 1), we perform replica exchange DMD simulations to efficiently sample the folding landscape of the polypeptides [56–60]. In each simulation, we start from a completely stretched conformation. We discard the first 0.5% of the trajectories, which have drastic energy and structural changes, to disregard the initial equilibration; we only use the equilibrated parts of the trajectories for analysis. First, we model XN1 (Fig. 1a) in its entirety to simultaneously study the interactions among all flanking regions, such as the Nt17, polyQ and polyP regions. Second, we model a mutant of XN1 that is missing the polyP regions: XN1-P<sub>11</sub>-P<sub>10</sub> (Fig. 1b) to study the structural effects of the polyP regions, which have been shown to protect against XN1



**Figure 1. Constructs studied.** Diagrams of the organization of the sequence regions in the (a) XN1 and (b) XN1-P<sub>11</sub>-P<sub>10</sub> constructs are shown. The number of residues in each region is indicated below the region. For all three constructs (XN1, XN1-P<sub>11</sub>-P<sub>10</sub> and the homopolymer  $Q_n$ , not shown) we vary the number of repeats modeled:  $n = 23, 36, 40, 47$ . In (a) XN1, the first N-terminal 17 residues are collectively referred to as the Nt17 region. Following Nt17 is the polyQ region, which contains a variable  $n$  number of glutamine repeats. P<sub>11</sub> and P<sub>10</sub> are the regions of 11 and 10 proline repeats respectively; they are referred to as the polyP regions. A region of 17 residues tethers the polyP regions together. Finally, there are 12 residues in the C-terminus of XN1. The (b) XN1-P<sub>11</sub>-P<sub>10</sub> construct is identical to XN1 with the exception that it does not contain the polyP regions. The XN1 sequence is explicitly written in Fig. 4d. We use the title of “construct” to refer to either XN1, XN1-P<sub>11</sub>-P<sub>10</sub> or  $Q_n$ . A “model” is a specific polypeptide, such as XN1Q<sub>23</sub>, which is a polypeptide of XN1 with 23 glutamine repeats. Thus a total of 12 models were studied, divided into 3 constructs with 4 different glutamine lengths.

doi:10.1371/journal.pcbi.1000772.g001

aggregation. By comparing the results from the XN1 models to the XN1-P<sub>11</sub>-P<sub>10</sub> models, we can determine the role of the polyP regions. Lastly, as controls, we model Q<sub>n</sub> homopolymers in the absence of all flanking regions.

### XN1Q<sub>n</sub> are less stable than Q<sub>n</sub> and XN1Q<sub>n</sub>-P<sub>11</sub>-P<sub>10</sub>

We apply the weighted histogram analysis method (WHAM) to analyze the folding thermodynamics of all simulated peptide systems [61]. For each peptide model, we calculate the heat capacity ( $C_V$ ) at different temperatures (Fig. 2). We find Q<sub>23</sub> undergoes a non-cooperative folding transition from an extended and unfolded state to a collapsed globule state, characterized by the broad and shallow  $C_V$  peak (Fig. 2a). Similar peaks, indicating a coiled-globule transition of a polyQ system, have been documented elsewhere [14]. As the length of Q<sub>n</sub> increases, the heat capacity peak gets taller and narrower, suggesting an increased folding cooperativity. In all three constructs (XN1, XN1-P<sub>11</sub>-P<sub>10</sub> and Q<sub>n</sub>) we find the transition temperature corresponding to the  $C_V$  peak (Fig. 2 and Table S1) is almost unaffected by the length of the polyQ region. Since the transition temperature is indicative of the polypeptide stability, our simulation suggests that the length of the polyQ region does not affect the stability of the polypeptides. Interestingly, we find that the average transition temperature for the XN1 models (311 K) is smaller than that of the Q<sub>n</sub> (322 K) and the XN1-P<sub>11</sub>-P<sub>10</sub> (340 K) models. The differences in transition temperatures indicate the XN1-P<sub>11</sub>-P<sub>10</sub> models are the most stable, followed by the Q<sub>n</sub> models and lastly, the XN1 models are the least stable.

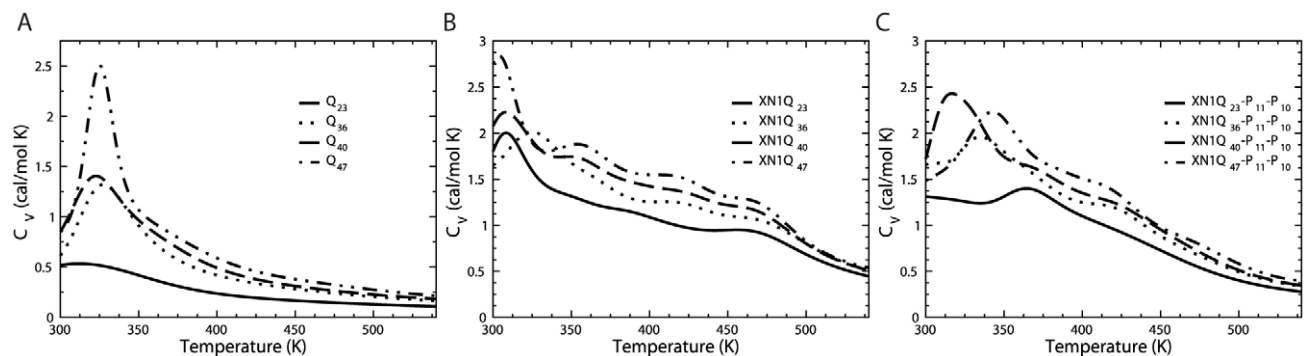
### Q<sub>n</sub> transitions from random coil to $\beta$ -sheet

As suggested by the Wetzel group and others [11,13,19], misfolding of polyQ monomers might initialize the aggregation and play a crucial role in the formation of amyloid fibrils. The peptide of polyQ is naturally unstructured, and thus, the length-dependent aggregation behavior can only be explained by the rare and spontaneous misfolding of the peptide. Therefore, we focus on the compact low-energy state from simulations (see Methods). The compact low-energy state usually constitutes only 13–27% of the total populations (Table S2). Among this subset of compact structures, we find a representative structure using a clustering algorithm (see Methods). These representative structures are not definitive misfolded states, but rather represent common, accessible compact states. We find that Q<sub>n</sub> is able to form a  $\beta$ -sheet for

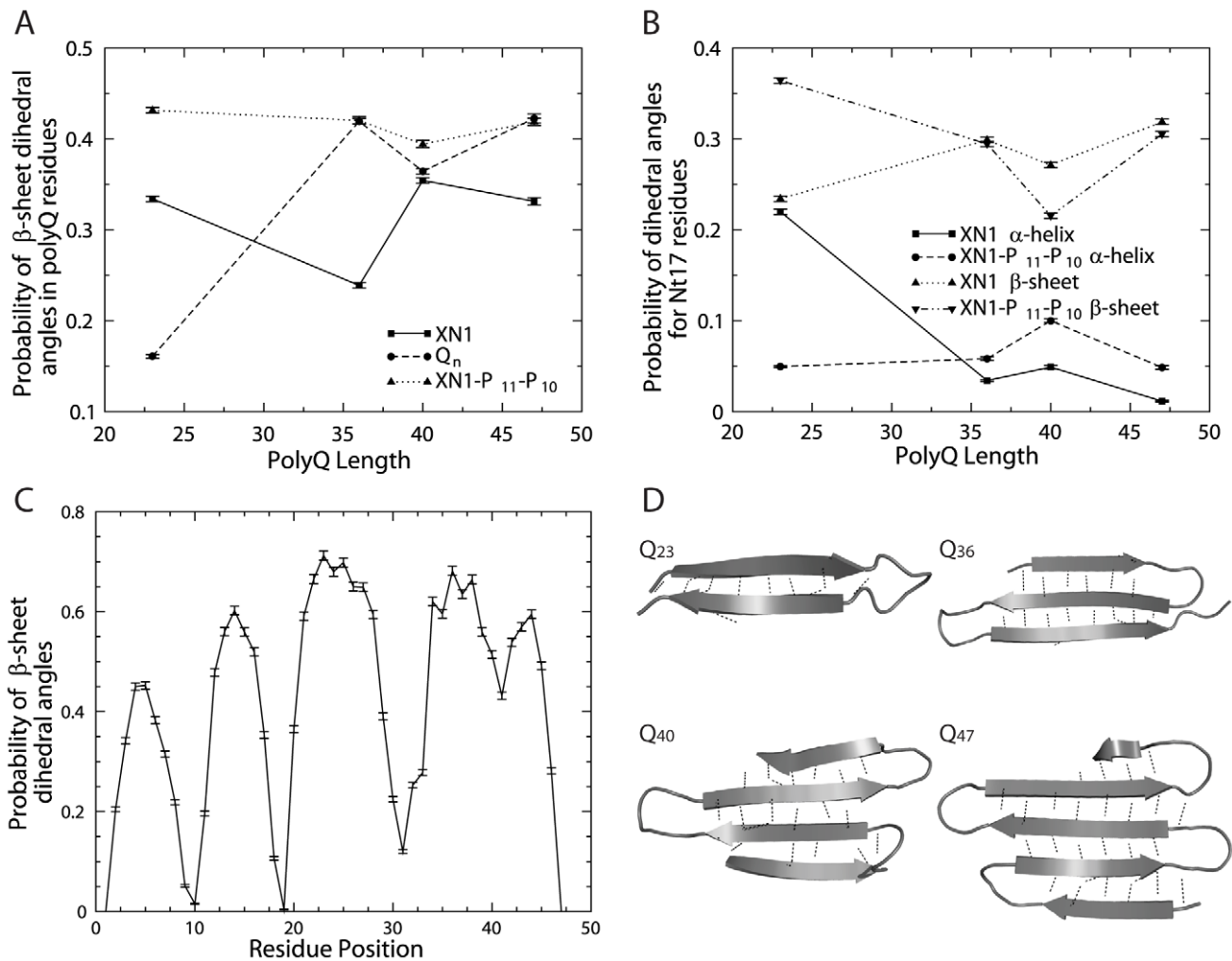
all lengths modeled (Fig. 3d). As the length of Q<sub>n</sub> increases, the  $\beta$ -sheet expands and contains more  $\beta$ -strands. Even the shortest polypeptide, Q<sub>23</sub>, is capable of forming a small  $\beta$ -sheet (a  $\beta$ -hairpin); although, this structure rarely forms during the simulation (Fig. 3d). Additionally, we calculate the probability of observing a given secondary structure for each residue (Fig. 3a,b,c) in the compact state (see Methods). For example, we find that in Q<sub>47</sub> there are many segments of the polypeptide featuring high  $\beta$ -conformation (Fig. 3c). Although the probability is computed for each residue, a high probability of  $\beta$ -conformation for consecutive residues indicates the formation of a  $\beta$ -strand (Fig. 3d). Furthermore, an overall perspective is gained when this probability is averaged over multiple residues (Figs. 3a and 3b). We find the compact states of Q<sub>23</sub> to be primarily unstructured; a typical residue in Q<sub>23</sub> is a random coil 67% of the time. In contrast, the residues in the long Q<sub>n</sub> models of Q<sub>36–47</sub>, have  $\beta$ -sheet dihedral angles 40% of the time. That is, residues in the long Q<sub>n</sub> models adopt  $\beta$ -sheet conformations almost 2.5 times more often than Q<sub>23</sub> (Fig. 3a). The relatively high  $\beta$ -sheet probabilities do not contradict the experimental observations of little to no monomeric  $\beta$ -sheet structure [19,15,25], since the secondary structure probability calculation is computed only for the small subset of compact and low-energy states. Additionally, the structural ensembles are constructed from the replica exchange simulations, which cover a wide-range of temperatures (see Methods). As a result, the distribution of conformations does not correspond to a single experimental condition, and thus, cannot be compared with experimental measurements. However, these ensembles can be used to evaluate the structural propensities of various polyQ constructs. From our calculations of the compact polypeptide, we find that Q<sub>n</sub> residues tend to transition from random coil conformations at short lengths, to  $\beta$ -sheet conformations at long lengths.

### PolyP hinders $\beta$ -sheet formation of polyQ in XN1

In the XN1 model, the polyQ region is surrounded by multiple flanking regions (Fig. 1), including the Nt17 and polyP regions. We find that for  $n = 36, 40$  and  $47$ , the residues in the polyQ region adopt  $\beta$ -sheet conformations approximately 10% less frequently than residues in isolated Q<sub>n</sub> (Fig. 3a). Thus, the flanking regions can lower the probability of the residues in the polyQ region to adopt  $\beta$ -sheet conformations. Based on secondary structure probabilities (data not shown) and the corresponding representa-



**Figure 2. Thermodynamics of the peptides.** Heat capacity and temperature curves are calculated with WHAM analysis of the simulation trajectories. (a) Calculations of the Q<sub>23</sub>, Q<sub>36</sub>, Q<sub>40</sub> and Q<sub>47</sub> models indicate the folding transition peaks become taller and narrower for longer glutamine lengths. (b) For the XN1Q<sub>23</sub>, XN1Q<sub>36</sub> and XN1Q<sub>47</sub> models, the transition temperatures are nearly identical (308K, 308K and 304K respectively). However, the XN1Q<sub>40</sub> model has a larger transition temperature around 325K. (c) The XN1Q<sub>23</sub>-P<sub>11</sub>-P<sub>10</sub>, XN1Q<sub>36</sub>-P<sub>11</sub>-P<sub>10</sub>, XN1Q<sub>40</sub>-P<sub>11</sub>-P<sub>10</sub> and XN1Q<sub>47</sub>-P<sub>11</sub>-P<sub>10</sub> models have varied transition peaks. The transition temperatures are 365K, 335K, 317K and 343K respectively. Detailed data on the peak positions are found in Table S1. doi:10.1371/journal.pcbi.1000772.g002



doi:10.1371/journal.pcbi.1000772.g003

tive structures (Figs. 4a, 4b), we find that the two polyP stretches in XN1 consistently form PPII helices (Fig. 4c). It is unlikely that the polyP regions fold into PPII helices due to interactions from neighboring regions. This is because the PPII helices are consistently found in every one of the compact structures; whereas, the other regions have more variable secondary structures. Thus, the polyP regions are likely forming PPII helices independently. We hypothesize that the P<sub>11</sub> and P<sub>10</sub> regions dominate the fold of

XN1, by forming these PPII helices, and subsequently, the remaining regions, including the polyQ region, are affected by these two PPII helices.

In order to study the effect of the PPII helices on the polyQ region, we perform simulations of XN1-P<sub>11</sub>-P<sub>10</sub> polypeptides that are sequentially identical to XN1 but lack the polyP regions (Fig. 1b). We find that residues in the polyQ region of XN1-P<sub>11</sub>-P<sub>10</sub> models adopt  $\beta$ -sheet conformations 10% more often than





## Discussion

### The length dependence of $Q_n$ aggregation

One intriguing phenomena of glutamine expansion diseases is the length dependence of disease onset [2,9]. It has been suggested that both short and long  $Q_n$  polypeptides can access similar misfolded structures that lead to aggregation, but the frequency at which this misfolded structure is visited depends on the length of  $Q_n$  [62]. That is, long  $Q_n$  aggregate fast, because they misfold frequently; contrarily, short  $Q_n$  rarely misfold and thus aggregate slow. Mounting experimental evidences support this model by showing that, regardless of length,  $Q_n$  polypeptides form aggregates of similar structure [25,27,28,39,62], which suggests the misfolded structures for all  $Q_n$  are also similar. Additionally, kinetic studies [16,21–27] verify the correlation between the length of the  $Q_n$  polypeptide and the rate at which it aggregates in solution. The remaining question is to determine the common misfolded structure that leads to aggregation. To this end, some investigators [8,11,16–18] have suggested that early forms of  $Q_n$  aggregates have high amounts of  $\beta$ -sheets. In the previous investigation by Wetzel's group, [11] glutamine homopolymers were capped by flanking lysine residues ( $K_2Q_nK_2$ ) to increase the peptide solubility [63]. It has been argued that the electrostatic repulsion between flanking lysines might prevent the formation of compact structures and alter the aggregation kinetics of the peptide system [19]. As the length increases, the screening effect reduces. Hence, despite the screening effect, the observation of intrinsic  $\beta$ -sheets formation of polyQ peptides is still valid. Our resulting extended  $\beta$ -sheet structures are prone to aggregation with exposed hydrogen bond donors and acceptors found in the polypeptide backbone [64]. By seeking to satisfy these bonds, the polypeptides can form bonds with other polypeptides that similarly have an exposed backbone, leading to the formation of large aggregates.

Our simulations of  $Q_n$  support a model, similar to one outlined by [27], wherein a common, compact misfolded state is accessible to most  $Q_n$  monomers and rates of misfolding are length-dependent. Accordingly, we find that partially folded monomers of both short and long  $Q_n$  polypeptides can have high amounts of  $\beta$ -character (Fig. 3d). Additionally, we find that the residues in long  $Q_n$  models, with 36 or more repeats, are more likely to have  $\beta$ -sheet conformations than short models,  $Q_{23}$  (Fig. 3a). The secondary structure probability of each individual residue is proportional to the probability of the overall polypeptide adopting that secondary structure. Thus we find a positive correlation between the length of the  $Q_n$  polypeptide and its probability of forming a  $\beta$ -sheet. Because previous studies [8,11,16–18] have linked  $\beta$ -sheet formation to aggregation, we suggest that long  $Q_n$  polypeptides are therefore more likely to form aggregates.

### The polyP regions protect XN1 from aggregation

Recent studies [1,3,5,20,39,40,42–44] indicate that the addition of a polyP region can inhibit aggregation of the polyQ region; this inhibition has been associated [40] with a PPII helix structure in the polyP region. We are able to find a structural effect on the polyQ region from the formation of PPII helices in the polyP regions. We find that the polyP regions form PPII helices and suppress the probability of polyQ residues in XN1 to adopt  $\beta$ -sheet dihedral angles. This suppression is still present for long polyQ lengths that are associated with disease. Upon removing these PPII helices, we find the polyQ residues are more likely to adopt  $\beta$ -sheet conformations; the likelihood is similar to that of the isolated glutamine homopolymers:  $Q_n$  (Fig. 3a). This similarity indicates that the other flanking regions of XN1 have little effect on the

probability of the polyQ region to form a  $\beta$ -strand. Furthermore, by considering that  $\beta$ -sheet formation has been linked to aggregation, we find that our results reflect experimental results. That is, because the polyP regions decrease the probability of polyQ residues adopting  $\beta$ -strand conformations, these polypeptides have a slower rate of aggregation, which is seen in other experiments [1,3,5,20,39,40,42–44]. Additionally, the polyP regions greatly destabilize XN1 polypeptides (Fig. 2b and 2c). Since unfolded polypeptides in the random coil state are unlikely to organize as an aggregate, the presence of the polyP in the XN1 sequence prevent it from folding into the aggregation-prone state. Therefore, we hypothesize that polyP regions protect XN1 from aggregation in two ways: 1) destabilize the polypeptide, and 2) the PPII helices formed by polyP inhibit formation of  $\beta$ -sheets.

### Nt17 misfolding in monomeric XN1

Currently, the role of the flanking regions in XN1 and other polyglutamine diseases is under debate [1,3,20,45–48]. One model [3,47] suggests that aggregation is initiated by the flanking Nt17 region. That is, initially, these flanking regions misfold and form oligomers; subsequently, there is an increase in the local concentration of the polyQ region, which causes the polyQ regions to misfold into protofibrils and ultimately mature, fatal fibrils. Others, however, contend that the native structure of the flanking regions is one that resists aggregation [1,46,48]. In a recent study [20], the expansion of polyQ repeats in XN1 is found to promote the misfolding of Nt17, which leads to rapid formation of oligomers. In particular, the structure of the Nt17 flanking region in XN1 is one part of this debate, which we discuss here.

We find a sharp decrease in the probability of the Nt17 residues to adopt  $\alpha$ -helix conformations as the length of the polyQ region increased. Concurrently, however, these residues are more likely to have  $\beta$ -strand conformations for longer polyQ repeat lengths (Fig. 3b). Hence, our results indicate misfolding of the Nt17 region; that is, the  $\alpha$ -helical native structure misfolds into a  $\beta$ -strand in the pathogenic associated XN1 models. Our simulation of the XN1 $Q_{47}$  suggests that both the Nt17 and polyQ regions simultaneously form  $\beta$ -strands (Fig. 4b). A possible future direction would be to identify the temperature at which the Nt17 residues transition from  $\alpha$ -helical to  $\beta$ -sheet conformation; a similar calculation has been done elsewhere [65]. In terms of the problem of the aggregation mechanism of XN1, the next step is to determine the role these two regions play in oligomerization of XN1. Here, our computational study suggests that the polyQ region also plays a critical role in the early stages of aggregation [20].

## Methods

### Discrete molecular dynamics

Unlike traditional molecular dynamic simulations, we discretize the spherically symmetric, pair-wise interaction potential in a DMD simulation [49–52], where, the continuous potential between any two atoms is reduced to a series of square well potentials. Such a simplification considerably accelerates the computation time because in square well potentials, the particles do not experience any force except at the boundaries of the square wells. Thus, the particles travel with constant momenta until a boundary is reached; at such a boundary, the two particles experience a force and are considered colliding. At a collision, the momenta, angular momenta and energies of only the two colliding particles are updated according to conservation laws. Thus, the most computationally intense process is to sort the event list to determine the next collision. Further details on the particular

version of DMD used here, such as interaction strength between atoms, are provided in reference [52]. Each polypeptide begins in an extended conformation; to allow for equilibration, the first 500 time units wherein the polypeptide has drastic energy and structural changes are discarded. We simulate each polypeptide in a cube with periodic boundaries. The dimension of the box is chosen to be large enough to fit the extended polypeptide.

#### All-atom models of Q<sub>n</sub>, XN1 and XN1-P<sub>11</sub>-P<sub>10</sub>

We model each polypeptide using an united all-atom approach, which is explained in detail elsewhere [52]. Briefly, this approach models all heavy atoms and polar hydrogen atoms in a polypeptide; interactions between atoms are governed by the Medusa Force Field [66]. The interactions include van der Waals (VDW) based on CHARMM19, orientation-dependent hydrogen bonding and implicit solvation EEF1 [67]. Because electrostatic interactions at long distances are weakened due to solvent screening, we currently do not model these effects in the all-atom DMD. Salt bridges between side chains are captured partially through the hydrogen bonding potential [52]. Despite the approximation of electrostatic interactions, we were able to fold six small proteins to their native state *ab initio* [52]. The sequence used to model XN1 is taken as the first 90 residues in Human Huntingtin Protein from NCBI [68] (Fig. 4d). There are charged, polar and non-polar groups scattered throughout the sequence. Additionally, the particular sequence, taken from NCBI, contains 23 glutamine repeats, which is a non-pathogenic length. The sequence is not modified except to add glutamines in the polyQ region as indicated or to remove the P<sub>11</sub> and P<sub>10</sub> stretches in the XN1-P<sub>11</sub>-P<sub>10</sub> models.

#### Replica exchange DMD simulations

To efficiently sample protein conformations, we use the replica exchange simulation technique [56–60]. With this technique, we are able to utilize multiple, parallel simulations of identical systems called replicas. For each of the 12 polypeptides, we perform simulations on eight replicas with the following set of temperatures: {0.85, 0.75, 0.68, 0.64, 0.6, 0.57, 0.53, 0.5}. The temperature units are in kcal/mol/k<sub>B</sub>, or about 500K. At a regular time interval of 500 time units (approximately 25 ps), we consider exchanging the temperature of two replicas. We only allow an exchange for two replicas with neighboring temperatures, for example 0.6 and 0.64. We use a Monte-Carlo based approach to accept or reject an exchange. The simulation length of each replica is  $1 \times 10^6$  time units (~50 ns).

#### Structure screening

Replica exchange simulations allow us to efficiently sample the conformational space of XN1 and its variants by simulating a wide range of temperatures. However, we focus only on the compact structures of the polypeptides. To screen for these compact states, we eliminate highly extended structures, which are those with a large radius of gyration ( $R_g$ ), and we include only those with low energy. The former criterion eliminates the transient structures explored during the early stages of folding when the polypeptide is far from its favored structure. The latter criterion selects for structures further along the folding pathway, because polypeptides lose energy during folding. We determine both the  $R_g$  cutoff (Fig. S1) and energy cutoff (Fig. S2) from a histogram of conformations sampled during the simulation. Each simulation produces 800,000 structures (1 conformation per 10 time units per replica). From this entire set, a subset of roughly 100,000–200,000 structures is selected through this screening process (Table S2). The energy cutoff is chosen to select for the lowest energy Gaussians.

#### Secondary structure probabilities

The probabilities calculated here are averages over the compact ensemble, which is a small subset of the entire population (Table S2). Thus, the calculations do not describe the polypeptides in general. Instead, the secondary structure likelihoods describe the polypeptides in a partially folded state, which estimates the misfolded structure. For each of the compact structures, we can calculate the backbone dihedral angles ( $\phi$  &  $\psi$ ) and the corresponding secondary structure for each residue. Then from the ensemble of compact structures for each polypeptide model, we compute the probability of a given residue to adopt  $\alpha$ -helix,  $\beta$ -sheet, turn or random coil dihedral angles (Fig. 3c). Furthermore, we also determine the secondary structure probability of a set of residues, or a region, by averaging the secondary structure probability over those residues (Figs. 3a, 3b). For example, a  $\beta$ -strand probability of 0.3 in the polyQ region means that on average, a given residue in the polyQ region has a 30% chance of adopting  $\beta$ -strand dihedral angles. These probabilities are calculated for individual amino acids regardless of neighboring residues. However, consecutive residues with high amount of calculated secondary structure probability will suggest the probability of forming specific secondary structures. Thus, this analysis method allows for comparison of secondary structure tendencies for polypeptides based on the compact ensemble. To calculate the error, we compute the standard error for each value; where the number of events is the number of times the energy of the trajectory crossed the energy cutoff (Fig. S3). In effect, this method counts the number of times during the folding trajectory that the polypeptide enters or exits the compact domain of its folding landscape.

#### Clustering

For visualization purposes, we identify representative structures for each polypeptide using an hierarchical clustering algorithm [69]. The structures used for clustering are taken from the simulation trajectories and are separated by at least 50ps. For most of the models studied, roughly 1500–2000 structures are used for clustering. As exceptions, the XN1Q<sub>23</sub> and Q<sub>23</sub> models include only about 1030 structures. The number of clusters is different for each polypeptide model and varied from 101–843. Similarly, the population of the largest cluster also varied. A detailed summary of these values for each polypeptide model is presented in the supplemental materials (Table S3). In this clustering scheme, the nodes are structures and the distance between a pair of nodes is the root mean square distance (RMSD) between the two structures. We use the single-linkage or minimum distance criterion for clustering. That is, the distance between a node and a cluster is equal to the distance between that node and the closest node in the cluster. The cutoff distance determines the size of clusters. If the distance between a node and a cluster is less than the cutoff, we include the node in that cluster; otherwise, we exclude the node from that cluster. We use a RMSD cutoff of 2.5Å for clustering the XN1Q<sub>47</sub>-P<sub>11</sub>-P<sub>10</sub> polypeptide and a 2Å RMSD cutoff for all other polypeptides. These cutoffs are chosen to maintain high structural similarity among the nodes in a cluster (Fig. S4). We study the centroid of each cluster, which is the most representative node or the centermost node of a cluster. Furthermore, we select the centroids from the largest clusters to be overall representatives of their respective polypeptides. To gauge the significance or reliability of the centroid, we calculate the ratio of the number of structures present in its cluster and the total number of structures considered for clustering. Example calculations are reported in the captions of Figs. 3d and 4. Finally, necessary raw data is presented in Table S3.

## Supporting Information

**Table S1** Thermodynamic Peak Values. Data corresponding to the peaks (Fig. 2) of the heat capacity versus temperature curves for each polypeptide modeled. Column 2 contains the values from the position of the peaks, which is the temperature at which the transition from folded to unfolded occurs. Column 3 contains the corresponding heat capacities at the folding transitions. Increasing glutamine repeat length does not correlate significantly with the transition temperature. The XN1 and Qn models show higher heat capacities for longer glutamine repeats.  
Found at: doi:10.1371/journal.pcbi.1000772.s001 (0.03 MB DOC)

**Table S2** Compact Structure Populations. Some statistics on the populations of compact structures that are used for analysis. Each simulation produces 800,000 structures in total. A subset of compact, low-energy structures is selected for analysis. Here, the total number of these compact structures is given for each simulation. For comparison, the percentages out of the total 800,000 structures represented by the compact populations are also shown. Because the compact populations are only 30% or less of the total, the polypeptides are primarily extended and unstructured.  
Found at: doi:10.1371/journal.pcbi.1000772.s002 (0.03 MB DOC)

**Table S3** Clustering Data. Statistics describing the clustering results. For each polypeptide studied, the following values are shown: total number of structures used for clustering (column 2), number of clusters (column 3) and the population of the largest cluster (column 4). A Root Mean Square Deviation (RMSD) cutoff of 2Å is used for all polypeptides, except XN1Q47-P11-P10 where 2.5Å is used. Due to computational constraints, only about 1% of all compact structures (Table S2) are used for clustering. The structures presented in Figures 3d and 4 are centroids of the largest cluster.  
Found at: doi:10.1371/journal.pcbi.1000772.s003 (0.03 MB DOC)

**Figure S1** Example Rg Histogram. Distribution of the Radius of Gyration (Rg). This histogram is an example from one polypeptide (XN1Q23). The bin size is 1Å. Exactly 800,000 Rg values, one per 10 time units per replica, from the trajectories of all eight replicas are used to generate the distribution. The dashed line indicates the Rg cutoff value (20Å in this case), which is used to determine compact structures. Structures with a larger Rg than the cutoff, are considered extended or insufficiently compact and are not included in the compact ensemble used for further analysis. Note the sharp peak around 16Å; most polypeptides produced a similar sharp peak. This peak indicates that most conformations explored during the simulation are compact.  
Found at: doi:10.1371/journal.pcbi.1000772.s004 (0.07 MB DOC)

## References

1. Truant R, Atwal RS, Desmond C, Munsie L, Tran T (2008) Huntington's disease: revisiting the aggregation hypothesis in polyglutamine neurodegenerative diseases. *FEBS J* 275: 4252–62.
2. Ross CA (2002) Polyglutamine Pathogenesis: Emergence of Unifying Mechanisms for Huntington's Disease and Related Disorders. *Neuron* 35: 819–822.
3. Saunders HM, Bottomley SP (2009) Multi-domain misfolding: understanding the aggregation pathway of polyglutamine proteins. *Protein Eng Des Sel* 22: 447–51.
4. Ross CA, Poirier MA (2004) Protein aggregation and neurodegenerative disease. *Nat Med* 10 Suppl: S10–7.
5. Monsellier E, Chiti F (2007) Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep* 8: 737–42.
6. Davies SW, Turmaine M, Cozens BA, Difiglia M, Sharp AH, et al. (1997) Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell* 90: 537–548.
7. Mangiarini L, Sathasivam K, Seller M, Cozens B, Harper A, et al. (1996) Exon 1 of the HD Gene with an Expanded CAG Repeat Is Sufficient to Cause a Progressive Neurological Phenotype in Transgenic Mice. *Cell* 87: 493–506.
8. Nekooki-Machida Y, Kurosawa M, Nukina N, Ito K, Oda T, et al. (2009) Distinct conformations of in vitro and in vivo amyloids of huntingtin-exon1 show different cytotoxicity. *Proc Natl Acad Sci U S A* 106: 9679–84.
9. Bates GP (2005) History of genetic disease: the molecular genetics of Huntington disease - a history. *Nat Rev Genet* 6: 766–73.

**Figure S2** Example Energy Histogram. Example histogram of two replicas from one polypeptide (XN1Q23). The bin size is 1 cal/mol. The distribution for each replica is based on 1,000,000 data points: one value per time unit. Here, the data from two replicas are pooled; therefore this distribution is based on 2,000,000 data points. The dashed line indicates the energy cutoff value (−285 cal/mol); this cutoff applies to all replicas of this polypeptide. The cutoff is chosen such that only the lowest energy states are included as compact structures.  
Found at: doi:10.1371/journal.pcbi.1000772.s005 (0.08 MB DOC)

**Figure S3** Trajectories Sampling the Compact Domain. Two example trajectories that depict how the simulations explored states that are compact and partially folded. Both trajectories are from the XN1Q23 simulation, and each trajectory corresponds to the particular replica indicated in the legend. The energy cutoff value is indicated as the red line. This value is determined by a histogram (Fig. S2). All explored states that occur below the cutoff are considered compact and are included in the compact ensemble.  
Found at: doi:10.1371/journal.pcbi.1000772.s006 (0.13 MB DOC)

**Figure S4** Hierarchical Clustering. One example (from XN1Q23) of how the cutoff distance for clustering is determined. With hierarchical clustering, we are able to determine the number of clusters remaining for a given Root Mean Square Deviation (RMSD) cutoff. The cutoff value, indicated by the dashed line (2Å), is chosen to maximize the clustering of similar structures while avoiding clustering of unlike structures. Clustering of similar structures occurs in the steeply descending portion of the curve (between 1Å and 2Å). In this region clusters are close to each other or similar, because a very small increase in the cutoff distance joins two clusters. Clustering of structures that are less similar occurs in the tail end of the curve (roughly 3Å or longer). Here, the clusters are distant or dissimilar and a large increase in the cutoff distance is required to join two clusters. Thus, a suitable cutoff is often found after the most similar clusters have been joined (here, between 1Å–1.5Å) and before the distant clusters are joined (2.5Å and longer).  
Found at: doi:10.1371/journal.pcbi.1000772.s007 (0.08 MB DOC)

## Acknowledgments

We would like to acknowledge Drs. Shantanu Sharma and Adrian Serohijos for technical assistance.

## Author Contributions

Conceived and designed the experiments: FD NVD. Performed the experiments: VVL FD NVD. Analyzed the data: VVL FD NVD. Wrote the paper: VVL FD NVD.



10. Rubinshtein DC, Leggo J, Coles R, Almqvist E, Biancalana V, et al. (1996) Phenotypic characterization of individuals with 30–40 CAG repeats in the Huntington disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36–39 repeats. *Am J Hum Genet* 59: 16–22.
11. Chen S, Ferrone FA, Wetzel R (2002) Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc Natl Acad Sci U S A* 99: 11884–9.
12. Bhattacharyya AM, Thakur AK, Wetzel R (2005) Polyglutamine aggregation nucleation: thermodynamics of a highly unfavorable protein folding reaction. *Proc Natl Acad Sci U S A* 102: 15400–5.
13. Sugaya K, Matsubara S (2009) Nucleation of protein aggregation kinetics as a basis for genotype-phenotype correlations in polyglutamine diseases. *Mol Neurodegener* 4: 29.
14. Vitalis A, Wang X, Pappu RV (2008) Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization. *J Mol Biol* 384: 279–97.
15. Lee CC, Walters RH, Murphy RM (2007) Reconsidering the mechanism of polyglutamine peptide aggregation. *Biochemistry* 46: 12810–20.
16. Poirier MA, Jiang H, Ross CA (2005) A structure-based analysis of huntingtin mutant polyglutamine aggregation and toxicity: evidence for a compact beta-sheet structure. *Hum Mol Genet* 14: 765–74.
17. Poirier MA, Li H, Macosko J, Cai S, Amzel M, et al. (2002) Huntingtin spheroids and protofibrils as precursors in polyglutamine fibrilization. *J Biol Chem* 277: 41032–7.
18. Wacker JL, Zareic MH, Fong H, Sarikaya M, Muchowski PJ (2004) Hsp70 and Hsp40 attenuate formation of spherical and annular polyglutamine oligomers by partitioning monomer. *Nat Struct Mol Biol* 11: 1215–22.
19. Walters RH, Murphy RM (2009) Examining polyglutamine peptide length: a connection between collapsed conformations and increased aggregation. *J Mol Biol* 393: 978–92.
20. Thakur AK, Jayaraman M, Mishra R, Thakur M, Chellgren VM, et al. (2009) Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. *Nat Struct Mol Biol* 16: 380–9.
21. Glabe CG (2006) Common mechanisms of amyloid oligomer pathogenesis in degenerative disease. *Neurobiol Aging* 27: 570–5.
22. Perutz MF, Johnson T, Suzuki M, Finch JT (1994) Glutamine repeats as polar zippers: Their possible role in inherited neurodegenerative diseases. *Proc Natl Acad Sci U S A* 91: 5355–5358.
23. Slepko N, Bhattacharyya AM, Jackson GR, Steffan JS, Marsh JL, et al. (2006) Normal-repeat-length polyglutamine peptides accelerate aggregation nucleation and cytotoxicity of expanded polyglutamine proteins. *Proc Natl Acad Sci U S A* 103: 14367–72.
24. Burke MG, Woscholski R, Yaliraki SN (2003) Differential hydrophobicity drives self-assembly in Huntington's disease. *Proc Natl Acad Sci U S A* 100: 13928–33.
25. Chen S, Berthelie V, Yang W, Wetzel R (2001) Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity. *J Mol Biol* 311: 173–82.
26. Masino L, Kelly G, Leonard K, Trotter Y, Pastore A (2002) Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. *FEBS Lett* 513: 267–272.
27. Klein FA, Pastore A, Masino L, Zeder-Lutz G, Nierengarten H, et al. (2007) Pathogenic and non-pathogenic polyglutamine tracts have similar structural properties: towards a length-dependent toxicity gradient. *J Mol Biol* 371: 235–44.
28. Vitalis A, Lyle N, Pappu RV (2009) Thermodynamics of beta-sheet formation in polyglutamine. *Biophys J* 97: 303–11.
29. Khare SD, Ding F, Gwanmesia KN, Dokholyan NV (2005) Molecular origin of polyglutamine aggregation in neurodegenerative diseases. *PLoS Comput Biol* 1: 230–5.
30. Chopra M, Reddy AS, Abbott NL, de Pablo JJ (2008) Folding of polyglutamine chains. *J Chem Phys* 129: 135102.
31. Marchut AJ, Hall CK (2007) Effects of chain length on the aggregation of model polyglutamine peptides: molecular dynamics simulations. *Proteins* 66: 96–109.
32. Ogawa H, Nakano M, Watanabe H, Starikov EB, Rothstein SM, et al. (2008) Molecular dynamics simulation study on the structural stabilities of polyglutamine peptides. *Comput Biol Chem* 32: 102–10.
33. Zanut D, Gunasekaran K, Lesk AM, Nussinov R (2006) Computational study of the fibril organization of polyglutamine repeats reveals a common motif identified in beta-helices. *J Mol Biol* 358: 330–45.
34. Perutz MF, Finch JT, Berriman J, Lesk A (2002) Amyloid fibers are water-filled nanotubes. *Proc Natl Acad Sci U S A* 99: 5591–5.
35. Barton S, Jacak R, Khare SD, Ding F, Dokholyan NV (2007) The length dependence of the polyQ-mediated protein aggregation. *J Biol Chem* 282: 25487–92.
36. Ignatova Z, Thakur AK, Wetzel R, Gierasch LM (2007) In-cell aggregation of a polyglutamine-containing chimera is a multistep process initiated by the flanking sequence. *J Biol Chem* 282: 36736–43.
37. Perutz MF (1996) Glutamine repeats and inherited neurodegenerative diseases: molecular aspects. *Curr Opin Struct Biol* 6: 848–858.
38. Ignatova Z, Gierasch LM (2006) Extended polyglutamine tracts cause aggregation and structural perturbation of an adjacent beta barrel protein. *J Biol Chem* 281: 12959–67.
39. Bhattacharyya A, Thakur AK, Chellgren VM, Thiagarajan G, Williams AD, et al. (2006) Oligoproline effects on polyglutamine conformation and aggregation. *J Mol Biol* 355: 524–35.
40. Darnell G, Orgel JP, Pahl R, Meredith SC (2007) Flanking polyproline sequences inhibit beta-sheet structure in polyglutamine segments by inducing PPII-like helix structure. *J Mol Biol* 374: 688–704.
41. Kim MW, Chelliah Y, Kim SW, Otwinowski Z, Bezprozvanny I (2009) Secondary structure of Huntingtin amino-terminal region. *Structure (London, England : 1993)* 17: 1205–12.
42. Dehay B, Bertolotti A (2006) Critical role of the proline-rich region in Huntingtin for aggregation and cytotoxicity in yeast. *J Biol Chem* 281: 35608–15.
43. Duennwald ML, Jagadish S, Muchowski PJ, Lindquist S (2006) Flanking sequences profoundly alter polyglutamine toxicity in yeast. *Proc Natl Acad Sci U S A* 103: 11045–50.
44. Zuchner T, Brundin P (2008) Mutant huntingtin can paradoxically protect neurons from death. *Cell Death Differ* 15: 435–42.
45. Kelley NW, Huang X, Tam S, Spiess C, Frydman J, et al. (2009) The predicted structure of the headpiece of the Huntingtin protein and its implications on Huntingtin aggregation. *J Mol Biol* 388: 919–27.
46. Williamson T, Vitalis A, Crick S, Pappu R (2009) Modulation of polyglutamine conformations and dimer formation by the N-terminus of Huntingtin. *J Mol Biol* 396: 1295–1309.
47. Tam S, Spiess C, Auyeung W, Joachimiak L, Chen B, et al. (2009) The chaperonin TRiC blocks a huntingtin sequence element that promotes the conformational switch to aggregation. *Nat Struct Mol Biol* 16: 1279–85.
48. Atwal RS, Xia J, Pinchev D, Taylor J, Epan RM, et al. (2007) Huntingtin has a membrane association signal that can modulate huntingtin aggregation, nuclear entry and toxicity. *Hum Mol Genet* 16: 2600–15.
49. Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI (1998) Discrete molecular dynamics studies of the folding of a protein-like model. *Fold Des* 3: 577–587.
50. Sharma S, Ding F, Dokholyan NV (2008) Probing protein aggregation using discrete molecular dynamics. *Front Biosci* 13: 4795–4807.
51. Chen Y, Ding F, Nie H, Serohijos AW, Sharma S, et al. (2008) Protein folding: then and now. *Arch Biochem Biophys* 469: 4–19.
52. Ding F, Tsao D, Nie H, Dokholyan NV (2008) Ab Initio Folding of Proteins with All-Atom Discrete Molecular Dynamics. *Structure* 16: 1010–1018.
53. Dokholyan NV (2006) Studies of folding and misfolding using simplified models. *Curr Opin Struct Biol* 16: 79–85.
54. Ding F, Dokholyan NV (2008) Dynamical roles of metal ions and the disulfide bond in Cu, Zn superoxide dismutase folding and aggregation. *Proc Natl Acad Sci U S A* 105: 19696–701.
55. Duennwald ML, Jagadish S, Giorgini F, Muchowski PJ, Lindquist S (2006) A network of protein interactions determines polyglutamine toxicity. *Proc Natl Acad Sci U S A* 103: 11051–6.
56. Hansmann UH (2002) Generalized ensemble techniques and protein folding simulations. *Comput Phys Commun* 147: 604–607.
57. Hansmann UH (2002) Protein-folding simulations in generalized ensembles. *Int J Quantum Chem* 90: 1515–1521.
58. Rhee YM, Pande VS (2003) Multiplexed-Replica Exchange Molecular Dynamics Method for Protein Folding Simulation. *Biophys J* 84: 775–786.
59. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314: 141–151.
60. Zhou R, Berne BJ, Germain R (2001) The free energy landscape for beta hairpin folding in explicit water. *Proc Natl Acad Sci U S A* 98: 14931–6.
61. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 13: 1011–1021.
62. Rossetti G, Magistrato A, Pastore A, Persichetti F, Carloni P (2008) Structural Properties of Polyglutamine Aggregates Investigated via Molecular Dynamics Simulations. *J Phys Chem B* 112: 16843–16850.
63. Chen S, Berthelie V, Hamilton JB, O'Nuallai B, Wetzel R (2002) Amyloid-like Features of Polyglutamine Aggregates and Their Assembly Kinetics. *Biochemistry* 41: 7391–7399.
64. Ding F, Dokholyan NV, Buldyrev SV, Stanley H, Shakhnovich EI (2002) Molecular Dynamics Simulation of the SH3 Domain Aggregation Suggests a Generic Amyloidogenesis Mechanism. *J Mol Biol* 324: 851–857.
65. Ding F, LaRocque JJ, Dokholyan NV (2005) Direct observation of protein folding, aggregation, and a prion-like conformational conversion. *J Biol Chem* 280: 40235–40.
66. Ding F, Dokholyan NV (2006) Emergence of protein fold families through rational design. *PLoS Comput Biol* 2: e85.
67. Lazaridis T, Karplus M (1999) Effective Energy Function for Proteins in Solution. *Proteins* 35: 133–152.
68. Human Huntingtin Protein. NCBI Reference Sequence: NP\_002102.4.
69. OC a cluster analysis program, Barton, G. J. 1993 (version 2.1a).