

# Polyglutamine fibrillogenesis: The pathway unfolds

Christopher A. Ross\*<sup>†</sup>, Michelle A. Poirier\*, Erich E. Wanker<sup>‡</sup>, and Mario Amzel<sup>§</sup>

\*Division of Neurobiology, Department of Psychiatry, Department of Neuroscience and Program in Cellular and Molecular Medicine, and <sup>§</sup>Department of Biochemistry and Biophysics, Johns Hopkins University School of Medicine, Baltimore, MD 21287; and

<sup>‡</sup>Neuroproteomics Section, Max Delbrueck Center for Molecular Medicine, Robert-Roessle-Strasse 10, D-13125 Berlin Buch, Germany

Nine neurodegenerative diseases are caused by expanding CAG repeats coding for polyglutamine (polyGln) (1–4). These include Huntington's disease, dentatorubral and pallidolusian atrophy, several forms of spino-cerebellar ataxia, and spinal and bulbar muscular atrophy. Within the central nervous system, each disease has a distinctive pattern of degeneration, with considerable overlap among the diseases (5, 6). The genes containing CAG repeats show no homology to each other outside of the glutamine repeats, and most are genes of unknown function. Thus, speculation concerning pathogenesis has focused on the polyGln expansion itself.

For all of these diseases, there is a threshold of repeat length that causes disease. This threshold varies somewhat among the different diseases, but is generally in the range of 35–45 consecutive glutamines. In all polyGln diseases, the age of disease onset is strongly correlated with polyGln length, so that above the threshold, a longer repeat results in an earlier age of onset.

A pathological hallmark of these diseases is the aggregation of mutant polyGln protein, resulting in the formation of intranuclear inclusion bodies. In some of the diseases, inclusions have been observed in the cytoplasm, dendrites, and axonal processes. The inclusions are generally seen in affected areas of the brain (7, 8), though not limited to those neurons most likely to degenerate (9). Thus, whether inclusions are responsible for neurotoxicity has been controversial. Some studies have indicated a correlation between polyGln-containing inclusions and disease progression (10). However, in other studies, inclusion formation was dissociated from cytotoxicity (11, 12). In fact, inclusion formation may be, in part, a reflection of cellular protective mechanisms (13). Nevertheless, the inclusions are a useful marker for pathology and may provide clues to pathogenesis.

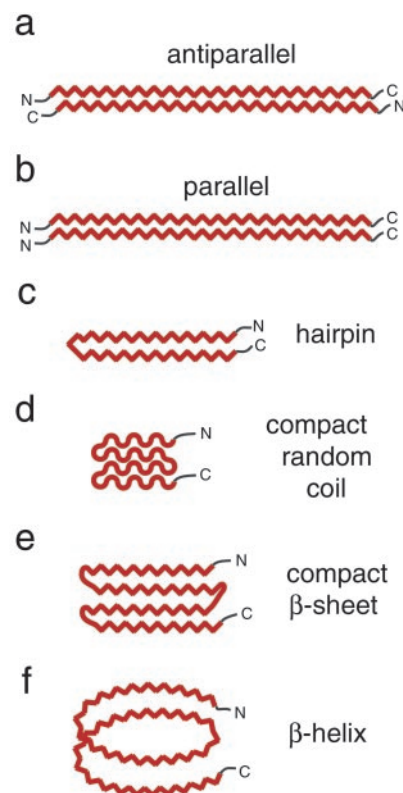
Aggregation of mutant polyGln proteins can be observed biochemically using a filter trap assay (14, 15). Aggregation *in vitro* proceeds by means of a nucleation-dependent process and results in the accumulation of  $\beta$ -sheet rich fibrillar structures detected by electron microscopy. Thus, the polyGln aggregation pathway appears to resemble that

of A $\beta$  protein in Alzheimer's disease and  $\alpha$ -synuclein in Parkinson's disease, as well as other amyloidogenic proteins (16–18). Even if polyGln inclusions are not the major toxic species, the aggregation process appears linked to pathogenesis. Therefore, it is critical to understand the structure of both normal and mutant polyGln stretches.

Detailed structural information on polyGln has been difficult to obtain, because both long and short stretches of synthetic polyGln peptides are quite insoluble. Nearly a decade ago, Max Perutz attempted to address this issue by using a Q<sub>15</sub> peptide flanked by basic residues to improve its solubility. He found this peptide to adopt  $\beta$ -structure, and constructed an atomic model of poly(L-glutamine) consisting of antiparallel  $\beta$ -sheets held together by hydrogen bonds between main-chain and side-chain amides (Fig. 1a). This structure, described as a "polar zipper," has been influential for studies of polyGln aggregation (14, 19, 20). Computer modeling studies have generated additional possible structures for expanded polyGln, such as parallel  $\beta$ -sheets (21),  $\beta$ -hairpins, and highly compact random coil (22) or  $\beta$ -sheet structures (Fig. 1b–e). Based on x-ray diffraction and electron microscopy data, Perutz and his group (23, 24) have recently suggested a polyGln  $\beta$ -helix model with 20 residues per turn (Fig. 1f).

Biophysical analysis of synthetic or recombinant polyGln peptides with stretches containing 5–44 consecutive glutamines have demonstrated that monomeric polyGln is unstructured (25–27). In contrast, expanded polyGln aggregates derived from isolated polyGln peptides or from recombinant polyGln-containing proteins adopt  $\beta$ -sheet structure, as shown by x-ray fiber diffraction studies, circular dichroism, Fourier transform infrared spectroscopy, and other methods (20, 25, 28, 29). Thus, it is likely that expanded polyGln sequences in the aggregated state are  $\beta$ -sheets, though detailed structural information is currently not available.

In a recent issue of PNAS, Thakur and Wetzel (30) provide a mutational analysis to address the question of polyGln aggregate structure. Previously, the Wetzel laboratory demonstrated that polyGln peptides have aggregation properties similar to exon-1 huntingtin (25).

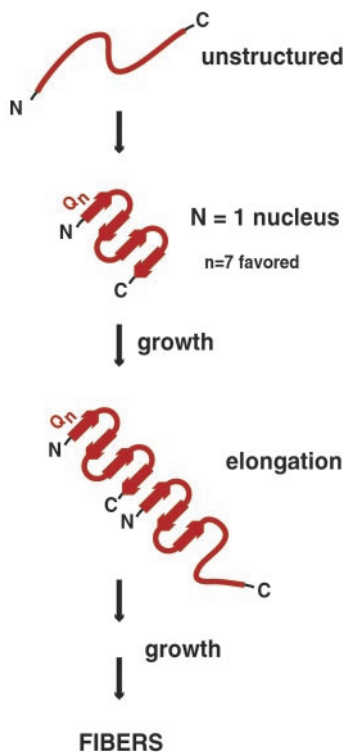


**Fig. 1.** Schematic representation of proposed structural models for aggregated mutant polyGln.  $\beta$ -Sheet is shown as a zig-zag. Expanded polyGln as an extended antiparallel  $\beta$ -sheet, first described by Max Perutz as a "polar zipper" (a), or as a parallel  $\beta$ -sheet (b). (c) An antiparallel  $\beta$ -hairpin comprised of two  $\beta$ -strands and a single  $\beta$ -turn. A highly compact structure, consisting of four antiparallel random coil (d) or  $\beta$ -strand (e) elements. (f) A parallel  $\beta$ -helix with 20 residues per turn. For simplicity, two polyGln molecules are shown each for a and b, whereas a single polyGln molecule is depicted in c–f.

In the current study, they used an ingenious strategy of inducing  $\beta$ -turns in synthetic peptides with Pro–Gly pairs at different intervals within a long polyGln stretch. To enhance the solubility of their peptides, they incorporated charged residues at both ends of the polyGln peptides, as introduced by Max Perutz and colleagues. Thakur and Wetzel investigated the influence of Pro–Gly

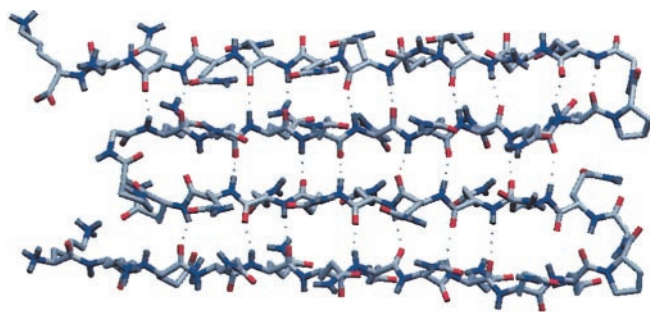
See companion article on page 17014 in issue 26 of volume 99.

<sup>†</sup>To whom correspondence should be addressed. E-mail: caross@jhu.edu.



**Fig. 2.** Model of polyGln aggregate initiation and elongation as proposed by Thakur and Wetzel (30). Before the conformational change that initiates disease pathogenesis, mutant polyGln lacks secondary structure. A polyGln monomer undergoes a structural transition to a four-stranded antiparallel  $\beta$ -sheet, with an optimum of seven glutamine residues per  $\beta$ -strand (extended chain). This structured monomer serves as a nucleus for binding of a second unstructured monomer. Binding of the disordered monomer to the ordered nucleus results in acquisition of  $\beta$ -structure in the newly added monomer, providing a new elongation site, and is referred to as template-assisted or "dock-and-lock" elongation. Adapted from Chen *et al.* (28).

mutations on formation of fibers, as detected by aggregation assays and electron microscopy, and found that peptides (ranging in total length from 46 to 50 aa) consisting of four  $Q_9$  or  $Q_{10}$  elements interspersed with three Pro-Gly

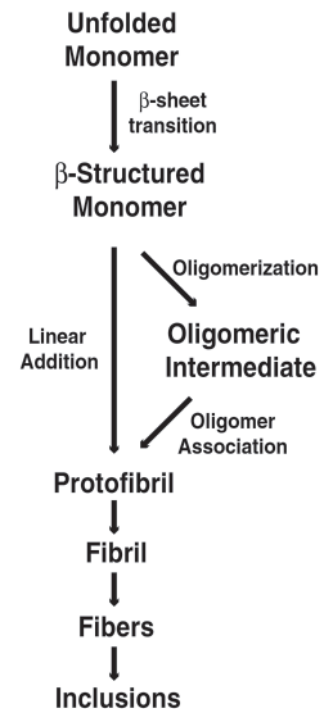


**Fig. 3.** Computer-generated drawing of  $PGQ_9$  as a four-stranded antiparallel  $\beta$ -sheet. Only main-chain-main-chain hydrogen bonds were built into the model. The stretches containing nine glutamine residues were built as antiparallel  $\beta$ -strands and the Pro-Gly pairs were built as turns. Atoms are colored as follows: carbons, red; nitrogens, blue; oxygens, red.

elements (termed  $PGQ_9$  or  $PGQ_{10}$ ) undergo spontaneous aggregation nearly as efficiently as  $Q_{45}$ .  $PGQ_7$  and  $PGQ_8$  peptides aggregate much less readily. Furthermore, a sequence containing D-prolines ( $P_DGQ_9$ ), which have a stronger preference for  $\beta$ -turns, aggregates more efficiently than the same peptide with L-prolines. By contrast, addition of a P residue in the center of a  $Q_9$  element blocked the peptide's ability to aggregate.

These data are suggestive of a model comprised of alternating  $\beta$ -strand and  $\beta$ -turn elements, with an optimum of seven consecutive glutamines within each  $\beta$ -strand (shown in Fig. 2). Based on electron microscopy data, the morphology of fibers generated with the  $PGQ_9$  peptide could be described as similar (though perhaps not identical) to that for the  $Q_{45}$  peptide. In addition, both the  $PGQ_9$  and the  $Q_{45}$  peptide were recognized by an anti-polyGln antibody selective for aggregates. Further, both peptides supported heterologous seeding (incorporation of normal length polyGln peptides into the expanded polyGln aggregate) in a similar manner. Thus, this study provides the first experimental data on the structure of aggregated polyGln, and is compatible with a model comprised of alternating  $\beta$ -strands and turns.

To further investigate the pathway of fiber assembly, Wetzel and colleagues recently conducted a kinetic analysis of polyGln aggregation at different concentrations of synthetic peptide (28). Their results, along with results of the present study, are summarized in Fig. 2. From their kinetic data, they conclude that aggregation of polyGln is initiated by a monomer that functions as the critical nucleus. They suggest that fibril formation proceeds via linear additions of single polyGln molecules, after nucleation events consisting of a random coil to  $\beta$ -sheet transition within an individual monomer. However, this analysis is indi-



**Fig. 4.** Hypothetical pathway of polyGln-mediated aggregation and inclusion formation. Unstructured polyGln monomer undergoes a structural conversion to  $\beta$ -sheet, resulting in the formation of protofibrillar intermediates. This step may proceed through a linear growth mechanism or through assembly of oligomeric intermediates. Protofibril assembly is followed by fibril formation, resulting in the characteristic inclusions observed in polyGln diseases and other amyloid-like diseases.

rect. An alternative pathway involving oligomeric intermediates has been suggested by biochemical and morphological studies of polyGln aggregation in the context of huntingtin exon-1 protein (29).

The present study provides the first experimental interventional data on the structure of polyGln within aggregates. It strongly supports a compact  $\beta$ -structure, rather than an extended strand model. Importantly, this structure is compatible with the polar zipper model, as hydrogen bonds are between main chains atoms and do not involve the side chains. A computer-generated drawing of the compact  $\beta$ -structure, with Pro-Gly residues at each  $\beta$ -turn, is depicted in Fig. 3. The strategy of using prolines for the introduction of  $\beta$ -turns is likely to be useful in future studies. This model provides testable hypotheses that can be addressed using other biophysical techniques, such as atomic force microscopy, NMR spectroscopy, electron paramagnetic resonance, and Raman spectroscopy. However, one limitation of the present study is that all data were derived from synthetic polyGln. Similar

studies will need to be carried out with exon-1 huntingtin and constructs encoding fragments or full-length versions of other polyGln proteins.

At present, the mechanism of polyGln-mediated toxicity is not well understood. Mutant polyGln may assume an abnormal conformation, allowing it to associate with short polyGln stretches in other critical cellular proteins, rendering them inactive. Biochemical data from the Wetzel laboratory support this possibility (25). The transcription factor CREB binding protein has been proposed as a target (31–33), as has the normal allele of huntingtin itself (34). Alternatively, abnormal interactions with transcription factors could occur by mechanisms not directly involving the expanded polyGln region (35–38). Another interesting possible mecha-

nism of toxicity is inhibition of the proteasome by the aggregated form of mutant polyGln protein (39, 40).

A key issue is the relationship between polyGln aggregation and cellular toxicity. As described above, the role of aggregation in toxicity has been controversial. Recent studies from the Wetzel laboratory have indicated that delivery of aggregated polyGln into cell nuclei is toxic (41). However, the exact nature of the material in the cells is uncertain. For delivery into cells, the aggregates were first sonicated and filtered. Thus, the toxic material could potentially be soluble monomer, an intermediate in the aggregation pathway, such as a monomeric nucleus as proposed by Thakur and Wetzel (30), or an oligomeric intermediate (29), as well as a small aggregate.

The pathway of polyGln-mediated aggregation is beginning to emerge, but

is still poorly understood (Fig. 4). It is unclear whether aggregation proceeds via linear addition of single molecules or whether there are oligomeric intermediates. In addition, the stage in the aggregation pathway that causes cell death is unknown. Whatever the mechanism, a critical first step to understanding the pathogenesis associated with polyGln-based neurodegenerative diseases is elucidating the structural basis of polyGln aggregation. Mutational analysis can now be used to answer some of these questions. PolyGln aggregation has already been used as a target for therapeutic development (42–47). A better structural understanding aggregation of polyGln, as well as aggregation of other proteins, may lead to development of rational therapeutics for polyGln diseases and for other neurodegenerative and protein aggregation diseases (48).

- Fischbeck, K. H. (2001) *Brain Res. Bull.* **56**, 161–163.
- Huang, C. C., Faber, P. W., Persichetti, F., Mittal, V., Vonsattel, J. P., MacDonald, M. E. & Gusella, J. F. (1998) *Somatic Cell Mol. Genet.* **24**, 217–233.
- Zoghbi, H. Y. & Orr, H. T. (2000) *Annu. Rev. Neurosci.* **23**, 217–247.
- Ross, C. (2002) *Neuron* **35**, 819–822.
- Ross, C. A. (1995) *Neuron* **15**, 493–496.
- Orr, H. T. (2001) *Genes Dev.* **15**, 925–932.
- DiFiglia, M., Sapp, E., Chase, K. O., Davies, S. W., Bates, G. P., Vonsattel, J. P. & Aronin, N. (1997) *Science* **277**, 1990–1993.
- Becher, M. W., Kotzuk, J. A., Sharp, A. H., Davies, S. W., Bates, G. P., Price, D. L. & Ross, C. A. (1998) *Neurobiol. Dis.* **4**, 387–397.
- Kuemmerle, S., Gutekunst, C. A., Klein, A. M., Li, X. J., Li, S. H., Beal, M. F., Hersch, S. M. & Ferrante, R. J. (1999) *Ann. Neurol.* **46**, 842–849.
- Davies, S. W., Turmaine, M., Cozens, B. A., DiFiglia, M., Sharp, A. H., Ross, C. A., Scherzinger, E., Wanker, E. E., Mangiarini, L. & Bates, G. P. (1997) *Cell* **90**, 537–548.
- Saudou, F., Finkbeiner, S., Devys, D. & Greenberg, M. E. (1998) *Cell* **95**, 55–66.
- Cummings, C. J., Reinstein, E., Sun, Y., Antalffy, B., Jiang, Y., Ciechanover, A., Orr, H. T., Beaudet, A. L. & Zoghbi, H. Y. (1999) *Neuron* **24**, 879–892.
- Kopito, R. R. (2000) *Trends Cell Biol.* **10**, 524–530.
- Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G. P., Davies, S. W., Lerach, H. & Wanker, E. E. (1997) *Cell* **90**, 549–558.
- Scherzinger, E., Sittler, A., Schweiger, K., Heiser, V., Lurz, R., Hasenbank, R., Bates, G. P., Lehrach, H. & Wanker, E. E. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 4604–4609.
- Serpell, L. C., Berriman, J., Jakes, R., Goedert, M. & Crowther, R. A. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 4897–4902.
- Rochet, J. C. & Lansbury, P. T., Jr. (2000) *Curr. Opin. Struct. Biol.* **10**, 60–68.
- Serio, T. R., Cashikar, A. G., Kowal, A. S., Sawicki, G. J., Moslehi, J. J., Serpell, L., Arnsdorf, M. F. & Lindquist, S. L. (2000) *Science* **289**, 1317–1321.
- Perutz, M. (1994) in *Protein Science* (Cambridge Univ. Press, London), pp. 1629–1637.
- Perutz, M. F. (1999) *Brain Res. Bull.* **50**, 467.
- Lathrop, R. H., Casale, M., Tobias, D. J., Marsh, J. L. & Thompson, L. M. (1998) *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 105–114.
- Starikov, E. B., Lehrach, H. & Wanker, E. E. (1999) *J. Biomol. Struct. Dyn.* **17**, 409–427.
- Perutz, M. F., Pope, B. J., Owen, D., Wanker, E. E. & Scherzinger, E. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 5596–5600.
- Perutz, M. F., Finch, J. T., Berriman, J. & Lesk, A. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 5591–5595.
- Chen, S., Berthelie, V., Yang, W. & Wetzel, R. (2001) *J. Mol. Biol.* **311**, 173–182.
- Masino, L., Kelly, G., Leonard, K., Trotter, Y. & Pastore, A. (2002) *FEBS Lett.* **513**, 267–272.
- Altschuler, E. L., Hud, N. V., Mazrimas, J. A. & Rupp, B. (1997) *J. Pept. Res.* **50**, 73–75.
- Chen, S., Ferrone, F. A. & Wetzel, R. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 11884–11889.
- Poirier, M. A., Li, H., Macosko, J., Cai, S., Amzel, M. & Ross, C. A. (2002) *J. Biol. Chem.* **277**, 41032–41037.
- Thakur, A. K. & Wetzel, R. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 17014–17019.
- Nucifora, F. C., Jr., Sasaki, M., Peters, M. F., Huang, H., Cooper, J. K., Yamada, M., Takahashi, H., Tsuji, S., Troncoso, J., Dawson, V. L., et al. (2001) *Science* **291**, 2423–2428.
- McCampbell, A., Taylor, J. P., Taye, A. A., Robitschek, J., Li, M., Walcott, J., Merry, D., Chai, Y., Paulson, H., Sobue, G., et al. (2000) *Hum. Mol. Genet.* **9**, 2197–2202.
- McCampbell, A., Taye, A. A., Whitty, L., Penney, E., Steffan, J. S. & Fischbeck, K. H. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 15179–15184.
- Zuccato, C., Ciammola, A., Rigamonti, D., Leavitt, B. R., Goffredo, D., Conti, L., MacDonald, M. E., Friedlander, R. M., Silani, V., Hayden, M. R., et al. (2001) *Science* **293**, 493–498.
- Steffan, J. S., Kazantsev, A., Spasic-Boskovic, O., Greenwald, M., Zhu, Y. Z., Gohler, H., Wanker, E. E., Bates, G. P., Housman, D. E. & Thompson, L. M. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 6763–6768.
- Steffan, J. S., Bodai, L., Pallos, J., Poelman, M., McCampbell, A., Apostol, B. L., Kazantsev, A., Schmidt, E., Zhu, Y. Z., Greenwald, M., et al. (2001) *Nature* **413**, 739–743.
- Shimohata, T., Nakajima, T., Yamada, M., Uchida, C., Onodera, O., Naruse, S., Kimura, T., Koide, R., Nozaki, K., Sano, Y., et al. (2000) *Nat. Genet.* **26**, 29–36.
- Dunah, A. W., Jeong, H., Griffin, A., Kim, Y. M., Standaert, D. G., Hersch, S. M., Mouradian, M. M., Young, A. B., Tanese, N. & Krainc, D. (2002) *Science* **296**, 2238–2243.
- Bence, N. F., Sampat, R. M. & Kopito, R. R. (2001) *Science* **292**, 1552–1555.
- Waelter, S., Boeddrich, A., Lurz, R., Scherzinger, E., Lueder, G., Lehrach, H. & Wanker, E. E. (2001) *Mol. Biol. Cell* **12**, 1393–1407.
- Yang, W., Dunlap, J. R., Andrews, R. B. & Wetzel, R. (2002) *Hum. Mol. Genet.* **11**, 2905–2917.
- Heiser, V., Scherzinger, E., Boeddrich, A., Nordhoff, E., Lurz, R., Schugardt, N., Lehrach, H. & Wanker, E. E. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 6739–6744.
- Heiser, V., Engemann, S., Bröcker, W., Dunkel, I., Boeddrich, A., Waelter, S., Nordhoff, E., Lurz, R., Schugardt, N., Rautenberg, S., et al. (2002) *Proc. Natl. Acad. Sci. USA* **99** (Suppl. 4), 16400–16406.
- Kazantsev, A., Walker, H. A., Slepko, N., Bear, J. E., Preisinger, E., Steffan, J. S., Zhu, Y. Z., Gertler, F. B., Housman, D. E., Marsh, J. L., et al. (2002) *Nat. Genet.* **30**, 367–376.
- Khoshnan, A., Ko, J. & Patterson, P. H. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 1002–1007.
- Marsh, J. L., Walker, H., Theisen, H., Zhu, Y. Z., Fielder, T., Purcell, J. & Thompson, L. M. (2000) *Hum. Mol. Genet.* **9**, 13–25.
- Sacchettini, J. C. & Kelly, J. W. (2002) *Nat. Rev. Drug Discovery* **1**, 267–275.
- Bennett, M. J., Huey-Tubman, K. E., Herr, A. B., West, A. P., Jr., Ross, S. A. & Bjorkman, P. J. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 11634–11639.