Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 5355–5358, June 1994
Biochemistry

Glutamine repeats as polar zippers: Their possible role in inherited neurodegenerative diseases

MAX F. PERUTZ, TONY JOHNSON, MASASHI SUZUKI, AND JOHN T. FINCH

Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England

Contributed by Max F. Perutz, February 25, 1994

ABSTRACT **Four inherited neurodegenerative diseases are linked to abnormally expanded repeats of glutamine residues in the affected proteins. Molecular modeling followed by optical, electron, and x-ray diffraction studies of a synthetic poly(L-glutamine) shows that it forms β-sheets strongly held together by hydrogen bonds. Glutamine repeats may function as polar zippers, for example, by joining specific transcription factors bound to separate DNA segments. Their extension may cause disease either by increased, nonspecific affinity between such factors or by gradual precipitation of the affected proteins in neurons.**

Four inherited neurodegenerative diseases are linked to abnormally expanded repeats of glutamines near the N termini of the affected proteins (for reviews, see refs. 1 and 2). They are Huntington disease (HD); spinal and bulbar muscular atrophy (SBMA), also known as Kennedy disease; spinocerebellar ataxia type 1 (SCA1); and dentatorubral-pallidoluysian atrophy (DRPLA) (3–8). All four diseases become more severe and begin earlier the longer the glutamine repeats. The repeats tend to lengthen in successive generations of affected individuals, especially in male transmission. So far no molecular function has been proposed for the glutamine repeats. We suggest that they act as polar zippers, joining protein molecules together similarly to the way leucine zippers join the transcription factors c-Jun and c-Fos.

The gene for the HD protein contains an open reading frame for >3100 amino acid residues. Its amino acid sequence shows no homology with any known protein. The glutamine-rich segment starts at residue 18. In healthy individuals its length varies between 16 and 33 residues; in those afflicted by HD it may vary between 35 and >100 residues. It is followed by an almost continuous stretch of 29 prolines. Poly(L-proline) forms a rigid helix with threefold screw symmetry and an axial repeat of 3.1 Å per residue (9). Its position near the N terminus of the HD protein implies that the glutamine repeat is mounted at the end of either a 90-Å-long stalk or, if an intervening peptide with the sequence QAQPLLPQPQ divides it into two, at the end of two consecutive, >30-Å-long stalks: a bizarre structure of a kind not encountered before.

SBMA is linked to the expansion of a glutamine-rich segment in the androgen receptor. This is a transcription factor made of a 920-residue chain with one DNA-binding C-terminal domain. Starting at residue 58, healthy individuals' receptors have a sequence of 15–31 glutamines. In SBMA patients this is expanded to 40–62 (4, 5). An androgen receptor with an expanded glutamine repeat trans-activated an androgen-responsive reporter gene more weakly than a normal receptor (10). No complete amino acid sequence of the protein responsible for SCA1 has been published. The length of its glutamine repeat varies from 25 to 36 residues in normal individuals from 43 to 81 residues in SCA1 patients (6).

Structure of Poly(L-glutamine). What could be the structure of poly(L-glutamine)? Perutz *et al.* (11) found several proteins containing repetitive sequences of polar residues. Molecular modeling showed these to be capable of linking β-strands together into sheets or barrels by networks of hydrogen bonds between their main-chain amides and between their polar side chains. Perutz *et al.* called these sequences polar zippers. One such sequence was a continuous stretch of up to 65 glutamines in the female sterile homeotic protein of *Drosophila* (12). Fig. 1 shows a computer-generated, stereochemically satisfactory model of two pairs of antiparallel β-strands of poly(L-glutamine) linked together by hydrogen bonds between their main-chain and side-chain amides.

When the importance of glutamine repeats in human disease became apparent, we decided to test the validity of this model experimentally, but this was difficult because poly(L-glutamine) is insoluble in water. To make it soluble, we synthesized a peptide with the sequence Asp_2-Gln_{15}-Lys_2.* Filtration of the peptide through Beckman Ultraspherogel SEC 2000 showed two fractions, one corresponding to the molecular weight of a monomer and the other to a broad distribution of aggregates of molecular weights in the range of hundreds of thousands. We investigated the conformation of the peptide by ultraviolet circular dichroism (CD). To avoid interactions between COO^- and NH_3^+ groups, we measured the CD spectra in solutions at pH 3.0 and 2.0, where the carboxylates are protonated. Peptides forming α-helices, β-sheets, or random coils give the different CD spectra shown in Fig. 2*a*. Fig. 2*b* shows the CD spectra of our peptide in three different solvents to be of the β-sheet type, even though one of the solvents, trifluoroethanol, normally induces formation of α-helices. Molar residue ellipticity at 197 nm was independent of peptide concentration, suggesting that the monomers form hydrogen-bonded hairpins and that the aggregates are made of tightly linked β-sheets. The spectra were qualitatively the same at pH 7.0 and 3.0. At pH 7.0 the peptide gradually precipitated. When viewed under

Abbreviations: HD, Huntington disease; SBMA, spinal and bulbar muscular atrophy; SCA1, spinocerebellar ataxia type 1; DRPLA, dentatorubral-pallidoluysian atrophy.

*The peptide was prepared by continuous-flow solid-phase synthesis by fluorenylmethoxycarbonyl chemistry on an automated synthesizer (NovaSyn Crystal) employing standard protocols (13, 14). On completion of the synthesis, the free N terminus was acetylated with acetic anhydride (20 equivalents) for 30 min. The peptide was cleaved from the Rink amide linker with trifluoroacetic acid/phenol/triethylsilane (23 ml/1 g/1 ml for 500 mg of peptide-resin assembly) and purified to homogeneity by preparative HPLC (Vydac C_8 column) (13, 14). Analysis of the purified material showed the following. Amino acid analysis: expected, $Asp_2Glx_{15}Lys_2$; found, $Asp_{1.91}Glx_{15.86}Lys_{2.00}$. Analytical reversed-phase HPLC (Aquapore RP-300, C_8 column): retention time, 12.62 min (>98%). Capillary zone electrophoresis (Beckman P/ACE system 2050; 75 mm × 50 cm capillary; 100 mM phosphate buffer, pH 2.5; 30 kV, 30°C); retention time, 15.19 min (>98%). Electrospray mass spectrometry: expected for Ac-$(Asp)_2$-$(Gln)_{15}$-$(Lys)_2$-NH_2, 2467.54; found, 2467.25, 2490.85 $(M+Na)^+$, 2505.63 $(M+K)^+$.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

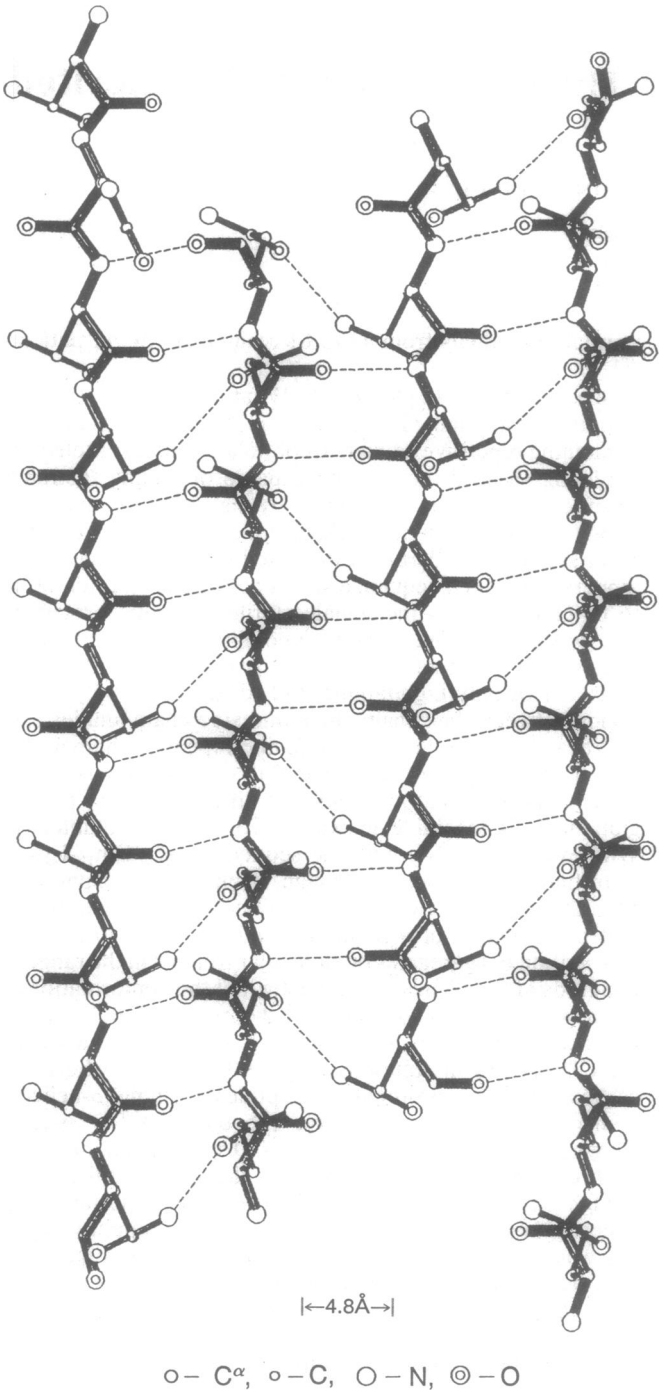

FIG. 1. Computer-generated structure of two paired antiparallel β-strands of poly(L-glutamine) linked together by hydrogen bonds between the main-chain and side-chain amides.

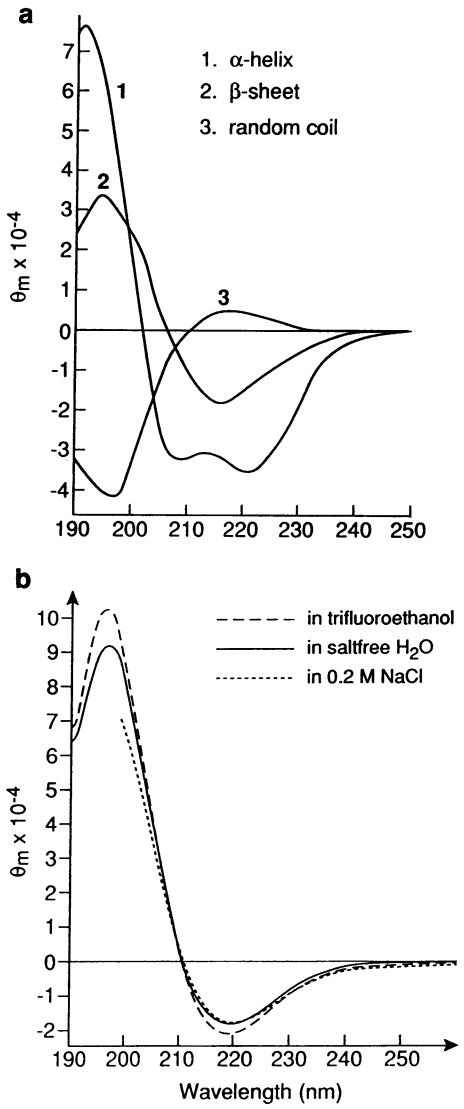

FIG. 2. (*a*) Standard CD spectra of α-helix, β-sheet, and random-coil structures of poly(L-lysine). (Reproduced by permission from ref. 15; copyright 1969, American Chemical Society.) (*b*) CD spectra of Asp_2-Gln_{15}-Lys_2 at 0.3 mg/ml in various media. The spectra were measured with a dichrograph (Jobin-Yvon CD6) at 20°C. Water and 0.2 M NaCl were adjusted to pH 3.0 by addition of trifluoroacetic acid. The pH of 95% 2,2,2-trifluoroethanol was 2.0.

the polarizing microscope, a suspension of the precipitate looked clear at first, but as the solvent between slide and coverslip began to evaporate, birefringence developed at the boundaries between air and water. The birefringence was positive, with the slow direction parallel to the meniscus. This suggested the presence of submicroscopic rod-shaped particles which surface tension had aligned parallel to the meniscus. We therefore examined the precipitate under the electron microscope. Fig. 3*a* shows that it consists of worm-like 70- to 120-Å-thick particles of varying length. An electron diffraction picture of a small clump of the dried particles (not shown) was dominated by a sharp ring of 4.8-Å spacing, diagnostic of the distance between neighboring polypeptide chains in β-sheets (16, 17). In addition, the picture contained faint rings at spacings of 8.4 and 4.2 Å. We next took an x-ray diffraction picture of the wet particles spun into a quartz capillary (not shown). This contained the same sharp rings as the electron diffraction picture plus another faint one at 2.8-Å spacing. A suspension of the particles was then dried on a glass slide. The dried film was lifted off and an x-ray diffraction picture was taken of it with the x-ray beam parallel to the plane of the film (Fig. 3). The picture exhibits a fiber diagram of the cross-β type with the 8.4-, 4.8-, and 4.2-Å reflections on the meridian and reflections at 3.6 and 3.2 Å on the equator. The 2.8-Å reflection lies beyond the rim of this picture. The equatorial reflection at 3.2 Å corresponds to the axial repeat per residue in a pleated β-sheet (16, 17). The reflection at 3.6 Å corresponds to the axial repeat per residue in a fully extended chain, which is not normally observed. Alternatively, it could represent a higher order of a long repeat that is not apparent in the rest of the picture. The strong meridional reflection at 4.8 Å and the equatorial one at 3.2 Å are indicative of a β-sheet with the chain direction

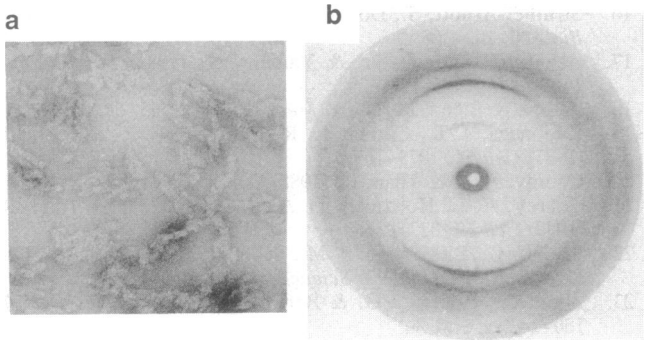

FIG. 3. (*a*) Electron micrograph of particles of Asp_2-Gln_{15}-Lys_2 formed in an aqueous solution at pH 7.0 and negatively stained with 1% uranyl acetate (× 90,000). (*b*) X-ray diffraction photograph (Cu K_α radiation; film distance, 37 mm) of the same particles oriented vertically, showing the dominant meridional reflection at 4.8 Å and the equatorial ones at 3.6 and 3.2 Å.

normal to the fiber axis. Molecular models show the spacing between successive β-sheets of poly(L-glutamine) hydrogen-bonded to each other by their amide side-chains to be about 17 Å, which suggests that the equatorial reflections at 8.4 and 4.2 Å and the diffraction ring at 2.8 Å could represent the second-, fourth-, and sixth-order reflections from parallel β-sheets stacked 16.8 Å apart. Odd orders may be weakened by the side-chain amides being stacked halfway between the main-chain amides. A likely structure suggested by the fiber diagram would be β-sheets rolled round each other. All these experimental results obtained from the 15-mer L-glutamine peptide are consistent with the polar zipper structure shown in Fig. 1.

Normal Functions of Glutamine Repeats. What possible function could the glutamine-rich segments in the HD, SBMA, SCA1, and SDRPLA proteins have? Long repeats of glutamines occur in many proteins, especially among transcription factors such as the homeobox proteins of *Drosophila*. Possible functions of glutamine repeats have been studied in the two homeoproteins coded for by the Abdominal B gene of *Drosophila*. Both proteins contain a C-terminal domain which includes the DNA-binding segment. One of them contains an additional N-terminal domain of 224 residues, 32% of which are glutamines. Deletion of part of this sequence did not diminish affinity for the homeoprotein-binding DNA segment, but it halved transcriptional activity. Conversely, splicing a distant part of the glutamine-rich domain to an otherwise inactive C-terminal domain restored transcriptional activity, which showed that the glutamine-rich domain did serve a necessary function, but without a hint as to its possible mechanism (18).

Does it serve a function in humans? Next to nothing is known as yet about the function of the glutamine repeats in the proteins affected by the three diseases, but some clues have emerged from studies of the human transcription factor Sp1. This is a trans-activator of gene expression which binds to G+C-rich regions in the promoters of several cellular and viral genes, including that of simian virus 40. The C-terminal fragment of Sp1 is a single chain of 696 residues with a glutamine-rich segment between residues 260 and 391 and with three zinc fingers beyond residue 540 (19). Courey and Tijan (20) designed an *in vivo* translation assay to study the function of the glutamine-rich segment. A C-terminal fragment of Sp1 containing the zinc fingers was transcriptionally inactive but became activated after a fragment containing residues 369–391 had been spliced onto it. This had the sequence PGNQVSWQTLQLQNLQVQNPQAQ; it contained 8 glutamines, 3 asparagines, a serine, and a threonine; these are all residues with polar side chains capable of forming hydrogen bonds with complementary polar side chains in neighboring β-strands (11). Alternatively, the inactive C-terminal fragment could be activated by splicing to it the glutamine-rich segment of the Antennapedia protein of *Drosophila* (21). The same authors next constructed a reporter plasmid with a single G·C box close to the initiation site of transcription, and six G·C boxes 1.7 kb downstream. Binding of Sp1 to these distantly placed G·C boxes stimulated the weak activation induced by the binding of Sp1 to the close G·C box. A similar stimulation was produced by the addition of a truncated Sp1 from which the zinc fingers had been deleted but which included the glutamine-rich segment (21). These experiments demonstrated that interaction between the glutamine-rich segments of Sp1 molecules bound to widely separated DNA segments enhanced transcription.

On the other hand, no function has so far been found for a stretch of 38 consecutive glutamines in the human TATA box-binding factor TFIID, but its conservation in the mouse factor argues in favor of its fulfilling an essential function (22, 23).

Molecular Pathology of Glutamine Repeats. Dominant transmission of HD and SCA1 argues in favor of these diseases being due to gains rather than losses of functions. SBMA is a chromosome X-linked recessive disease that manifests itself only in males. In females X-inactivation produces somatic cell mosaicism in which only the abnormal gene is likely to be expressed in a proportion of cells and only the normal gene in the remainder. The neural network is thought to contain enough redundancy so that inactivation of those neurons in which the abnormal gene is expressed does not matter. Loss of function is excluded as a cause of SBMA by the absence of SBMA symptoms in patients suffering from deletion of the androgen receptor gene. Mhatre *et al.* (10) have therefore suggested that the gain of function in each of the three diseases consists in an "aberrant transregulatory activity."

One possibility is that glutamine repeats joining specific complementary proteins as part of normal transcriptional regulation acquire excessively high affinities for each other or acquire nonspecific affinities for other regulatory proteins when they become too long. Alternatively, and perhaps more probably in the light of our results, extensions of their glutamine repeats may cause the affected proteins to agglomerate and precipitate in neurons; symptoms may set in when these precipitates have reached a critical size or have resulted in a critical number of neural blocks. This would explain better why symptoms appear earlier in life and become more severe the longer the extension of the glutamine repeats and why the main histological manifestation of HD consists in neural degeneration. On the other hand, immunostaining of Purkinje cells and cells in the human frontal cerebral cortex with an antibody against the HD protein showed no difference between cells from a normal individual and those from a HD patient (24).

1. Martin, J. B. (1993) *Science* **262,** 674–676.
2. Ross, A. C., McInnis, M. G., Margolis, R. L. & Li, S.-H. (1993) *Trends Neurosci.* **16,** 254–259.
3. The Huntington's Disease Collaborative Group (1993) *Cell* **72,** 971–983.
4. La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fischbeck, K. H. (1991) *Nature (London)* **352,** 77–79.
5. La Spada, A. R., Roling, D. B., Harding, E. A., Warner, C. L., Spiegel, R., Hausmanova-Petrusewicz, I., Yee, W.-C. & Fischbeck, K. H. (1992) *Nature Genet.* **2,** 301–304.
6. Orr, H. T., Chung, M.-y., Banfi, S., Kwiatkowski, Jr., T. J., Servadio, A., Beaudet, A. L., McCall, A. E., Duvick, L. A., Ranum, L. P. W. & Zoghbi, H. Y. (1993) *Nature Genet.* **4,** 221–226.
7. Matilla, T., Volpini, V., Genis, D., Rosell, J., Corral, J., Davalos, A., Molins, A. & Estivill, X. (1993) *Hum. Mol. Genet.* **2,** 2123–2128.
8. Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S.,

Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T., Saito, M., Tomoda, A., Miike, T., Naito, H., Ikuta, F. & Tsuji, S. (1994) *Nature Genet.* **6**, 9–13.
9. Cowan, P. M. & McGavin, S. (1955) *Nature (London)* **176**, 501–503.
10. Mhatre, A. N., Trifiro, M. A., Kaufman, M., Kazemi-Esfarani, P., Figlewicz, D., Rouleau, G. & Pinsky, L. (1993) *Nature Genet.* **5**, 184–188.
11. Perutz, M. F., Staden, R., Moens, L. & De Baere, I. (1993) *Curr. Biol.* **3**, 249–253.
12. Haynes, S. R., Mozer, B. A., Bhatin-Dey, N. & Dawid, I. B. (1989) *Dev. Biol.* **134**, 246–257.
13. Atherton, E. & Sheppard, R. C. (1988) *Solid Phase Peptide Synthesis: A Practical Approach* (Oxford Univ. Press, Oxford).
14. Dryland, A. & Sheppard, R. C. (1988) *Tetrahedron* **44**, 859.
15. Greenfield, N. & Fasman, G. D. (1969) *Biochemistry* **8**, 4108–4116.
16. Struther-Arnott, S., Dover, S. D. & Elliott, A. (1967) *J. Mol. Biol.* **30**, 202–208.
17. Ashida, T., Tanaka, I. & Yamane, T. (1981) *Int. J. Pept. Protein Res.* **17**, 322–329.
18. Ali, N. & Bienz, M. (1991) *Mech. Dev.* **35**, 55–64.
19. Kadonaga, J. T., Carner, K. R., Masiarz, F. R. & Tijan, R. (1987) *Cell* **51**, 1079–1090.
20. Courey, A. J. & Tijan, R. (1988) *Cell* **55**, 887–898.
21. Courey, A. J., Holtzman, D. A., Jackson, S. P. & Tijan, R. (1989) *Cell* **59**, 827–836.
22. Kao, C. C., Lieberman, P. M., Schmidt, M. C., Zhou, Q., Pei, R. & Berk, A. J. (1990) *Science* **248**, 1646–1650.
23. Zhou, Q., Boyer, J. G. & Berk, A. J. (1993) *Genes Dev.* **7**, 180–187.
24. Hoogeveen, A. T., Willemsen, R., Meyer, N., de Rooij, K. E., Roos, R. A. C., van Ommen, G.-J. B. & Galjaard, H. (1993) *Hum. Mol. Genet.* **2**, 2069–2073.