Opinion

GLUTAMINE REPEATS AS POLAR ZIPPERS: THEIR ROLE IN INHERITED NEURODEGENERATIVE DISEASE

Max F. Perutz
MRC Laboratory of Molecular Biology, Cambridge, United Kingdom

I want to recount how my interest in the control of the oxygen affinity of the heme iron by the globin in hemoglobin diverted me to the study of the molecular mechanism of hereditary neurodegenerative diseases.

It began with my curiosity about why the equilibrium constant of the reaction of hemoglobin with oxygen varies over five orders of magnitude in different species, even though the structure of the globin surrounding the heme remains unaltered. At the low-affinity end are the hemoglobins of teleost fish, which can discharge their oxygen into the swim bladder against high pressures. At the other end of the scale lies a hemoglobin of the parasitic nematode Ascaris, which has an oxygen affinity so high that its oxygen fails to dissociate even in a vacuum.

Ascaris hemoglobin is an octamer of MW 324,800. Each of its subunits is made up of two tandemly linked globin domains with an 18 residue C-terminal extension in which the sequence Glu-Glu-His-Lys is repeated four times. When this sequence is plotted on a β-strand, one side is lined with alternate glutamates and lysines, and the other with alternate glutamates and histidines. There have been many reports of leucine zippers, which tie α-helices together by mainly nonpolar interactions, but the Ascaris sequence suggested that it acted as a *polar* zipper, a feature not found in proteins before (1,2).

It seemed unlikely that this was unique in nature. A computer search of the protein database soon revealed a variety of other polar zipper sequences, among them proteins with long repeats of only glutamines. Suspecting that these could also form zippers, I built a molecular model which suggested that β-strands of poly-L-glutamine could aggregate into β-pleated sheets in which the strands are linked together by hydrogen bonds between their main chain and their side-chain amides (Fig. 1). I wrote a short paper on polar zippers in proteins and believed that this was the end of the story (3).

No sooner had this appeared, than I read the astonishing paper in *Cell* of March 1993 on the gene for Huntington's Disease (4), a dominantly inherited, severe neurodegenerative disease which usually sets in in middle age and leads first to uncontrolled movements, then to mood disturbances, depression, and death. The gene consisted of 67 exons spread over 180 kb of DNA; it contained an open reading frame for a protein of 3144 amino acid residues with a sequence devoid of homology with any other known proteins. Beginning at residue 18 was a long repeat of glutamines, followed by a long repeat of prolines. The only difference between the normal protein and that found in patients with Huntington's Disease (HD) consisted in the length of the glutamine repeat: below 35 in the normal protein and above 35 in the diseased ones. The severity of the disease was reported to increase, and the age of onset to decrease, with the number of glutamine repeats. In heterozygotes, both the normal and the abnormal protein are expressed in all tissues, but the abnormal one kills only neurons, specifically those in the striatum of the brain. Both proteins are found in he cytoplasm rather than the nucleus of neurons, and they are concentrated in the cell bodies, dendrites, fibers, and nerve endings (5–8).

Glutamine repeats and glutamine-rich domains also occur in many homeodomain proteins of Drosophila and in the human transcription factor SP1. They do not seem to have aroused the curiosity of protein chemists, but in vitro transcription experiments have shown that

|←—4.8Å—→|

o-C_α.   •-C.   ○-N.   ⊚-O
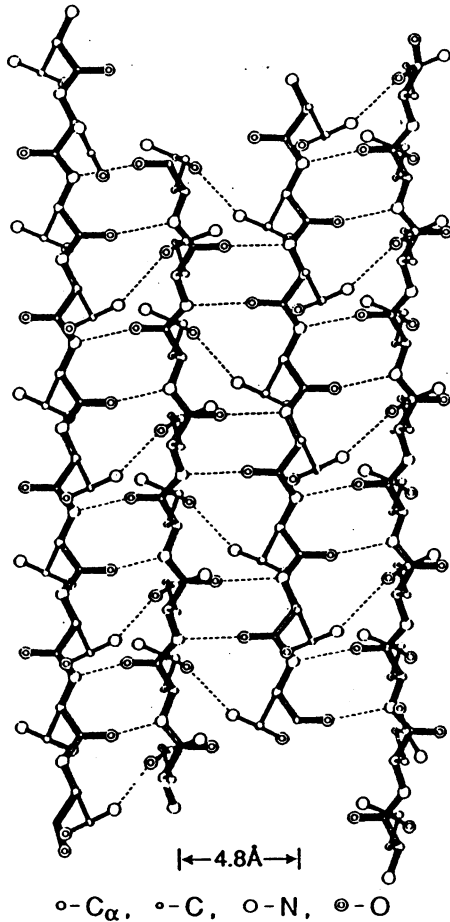
FIG. 1.   Computer-generated structure of two paired antiparallel β-strands of poly (L-glutamine) linked together by hydrogen bonds between the main-chain and side-chain amides

glutamine-rich domains promote interaction between transcription factors bound to distant promoters or between transcription factors and their associated factors. Thirty-three of forty proteins with at least 20 glutamine repeats were found to be transcription factors, but the HD protein cannot be a transcription factor, because it is absent in the nucleus. Its function is unknown (9–13).

Not only is the normal function of the HD protein unknown, but its glutamine repeat may not even have a normal function, because in the puffer fish (*Fugu rubripes*) it is reduced to four glutamines, probably too few for stable association with other proteins (14). Its expression in primates may be a genetic accident that conveyed no selective advantage.

It occurred to me that the polar zipper action of glutamine repeats might furnish a possible clue to the molecular mechanism of the disease,

but in view of the medical importance of the problem it seemed essential to test that idea experimentally, first with a synthetic polypeptide and then a protein. Poly-L-glutamine alone is insoluble in water. To render it soluble, my colleague Tony Johnson synthesized the polypeptide Asp_2-Gln_{15}-Lys_2. This formed large aggregates which circular dichroism, electron, and X-ray diffraction proved to consist of β-pleated sheets, consistent with my polar zipper idea, even though isolated glutamines in proteins tend to favor α-helix formation (15).

It remained to find out whether incorporation of glutamine repeats into a small monomeric protein of known structure would make it associate into oligomers. On the advice of Prof. A. R. Fersht, I chose chymotrypsin inhibitor–2 (CI2) from barley seeds, a small protein with an external loop of nine residues carrying the inhibitory methionine. Kelvin Stott, Jonathan Blackburn, and I engineered two constructs of CI2 from which the first 20 disordered residues had been deleted. In one construct, we replaced the loop by the sequence Gly-Gln_{10}-Gly, and in the other we inserted that sequence into the loop. The two constructs, together with the truncated wild-type protein, were expressed in good yield in *Escherichia coli*. Fractionation of the cell extracts by fast-protein liquid chromatography (FPLC) yielded the truncated wild type as a single fraction, while each of the constructs yielded two additional fractions of higher molecular weight. Electrospray mass spectrometry of the fractions gave molecular weights that agreed to within the nearest proton with those calculated for the truncated wild type and the two constructs as monomers. On the other hand, analytical ultracentrifugation showed that the wild-type fraction contained only monomers, whereas the additional fractions contained in the two constructs represented dimers and trimers. The constructs gave circular dichroism spectra similar to those of the truncated wild type, showing that its structure must have been preserved in the constructs. A circular dichroism difference spectrum of the trimer versus the truncated wild type was similar to the β-pleated sheet spectrum of Asp_2-Gln-Lys_2, while that of the monomer versus wild type resembled the spectrum of type I β-turns.

It seemed important to determine the dissociation constants of the CI2 oligomers, but they proved stable up to 70°C, above which they denatured. Nor did they dissociate in urea at concentrations below 3 M; at higher urea concentrations they denatured cooperatively with their

glutamine repeats. These findings implied that both association and dissociation rates are extremely slow. The slow association rates of the monomers may be due to hydrogen bonds between main chain and side chain amides within their glutamine repeats which are folded into hairpins, while the slow dissociation rates of the oligomers are clearly due to the many intermolecular hydrogen bonds between the glutamine repeats linked into β-pleated sheets (16).

What did these findings imply for the molecular mechanism of HD and the other four recently discovered inherited neurodegenerative diseases due to expanded glutamine repeats (17,18)? In HD, neurodegeneration is believed to result from a gain of an unknown toxic function rather than a loss of function (19). Our experiments show that glutamine repeats act as polar zippers whose expansion may lead to either aberrant self-association of the affected proteins or to their association with other proteins. The most obvious inference would be that they cause self-association into large protein aggregates, but immunostaining of the degenerate neurons in postmortems of patients has shown no evidence for these. Alternatively, association with other proteins may poison the cells.

Li et al. searched for proteins with which the rat and human HD proteins associate (20). They screened a rat brain DNA library by ligating a cDNA encoding the first 230 amino acids of the HD protein to a Gal-4 DNA binding domain plasmid vector of the yeast two hybrid system. They found one positive clone among 100,000 colonies. It had an insert of 600 base pairs encoding part of a new hydrophilic protein which associated with the human HD protein. Expansion of the glutamine repeat of the HD protein enhanced the association. The full protein of more than 600 amino acid residues contains a repeat of glutamates which would associate with glutamine repeats even more strongly than glutamine repeats would associated with themselves, because hydrogen bonds between amides and carboxylates are stronger than hydrogen bonds between amides. The protein is enriched also in human brain and appears low or absent from other tissues. Assuming that the human protein does the same as the rat protein, its discovery takes the search for the molecular mechanism of HD a big step further.

Even so, there remains a host of formidable questions to be answered. It is worth reflecting that I was led to the idea of glutamine repeats as polar zippers not because I had planned to work

on neurodegenerative disease, but because I was curious about the high oxygen affinity of the hemoglobin of *Ascaris lumbricoides*, the intestinal round worm that infects approximately a billion people in the world today. I have often found that in a strange and unpredictable way my research plans me.

REFERENCES

1. De Baere I, Liu L, Moens L, et al. (1992) Polar zipper sequence in the high affinity hemoglobin of *Ascaris suum:* Amino acid sequence and structural interpretation. *Proc. Natl. Acad. Sci. U.S.A.* **89:** 4638–4642.
2. Sherman DR, Kloek AP, Krishnan BR, Guinn B, Goldberg DE. (1992) *Ascaris* hemoglobin gene: Plant-like structure reflects the ancestral globin gene. *Proc. Natl. Acad. Sci. U.S.A.* **89:** 11696–11700.
3. Perutz MF, Staden R, Moens L, De Baere I. (1993) Polar zippers. *Curr. Biol.* **3:** 249–253.
4. The Huntington's Disease Collaborative Group. (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72:** 971–983.
5. Li S-H, Schilling G, Young III WS, et al. (1993) Huntington's disease gene (IT15) is widely expressed in human and rat tissues. *Neuron* **11:** 985.
6. Sharp AH, Loev SJ, Schilling G, et al. (1995) Widespread expression of Huntington's disease gene (IT15) protein product. *Neuron* **14:** 1065–1074.
7. Trottier Y, Devys D, Imbert G, et al. (1995) Cellular localisation of the Huntington's disease protein and discrimination of the normal and mutated form. *Nature Genet.* **10:** 104.
8. Persichetti F, Ambrose CM, Ge P, et al. (1995) Normal and expanded Huntington's disease alleles produce distinguishable proteins due to translation across the CAG repeat. *Mol. Med.* **1:** 374–383.
9. Ali N, Bienz M. (1991) Functional dissection of abdominal-β protein. *Mech. Dev.* **35:** 55–64.
10. Courey AJ, Tjian R. (1988) Analysis of SP1 *in vivo* reveals multiple transcriptional domains, including a novel glutamine-rich motif. *Cell* **55:** 887–898.
11. Courey AJ, Holtzman DA, Jackson SP, Tjian R. (1989) Synergistic activation by the glu-

tamine-rich domains of human transcription factor SP1. *Cell* **59:** 827–836.

12. Hoey TH, Weinziere ROJ, Gill G, Chen J-L, Dynlacht BD, Tjian R. (1993) Molecular cloning and functional analysis of *Drosophila* TAF110 reveal properties expected of coactivators. *Cell* **72:** 247–260.

13. Gerber HP, Seipel K, Georgiev O, et al. (1994) Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* **263:** 808–811.

14. Baxendale S, Abdulla S, Elgar G, et al. (1995) Comparative sequence analysis of the human and pufferfish Huntington's disease genes. *Nature Genet.* **10:** 67–76.

15. Perutz MF, Johnson T, Suzuki M, Finch JT. (1994) Glutamine repeats as polar zippers: Their possible role in inherited neurodegeneative diseases. *Proc. Natl. Acad. Sci. U.S.A.* **91:** 5355–5358.

16. Stott K, Blackburn JM, Butler PJG, Perutz M. (1995) Incorporation of glutamine repeats makes protein oligomerize: Implications for neurodegenerative diseases. *Proc. Natl. Acad. Sci. U.S.A.* **92:** 6509–6513.

17. La Spada AR, Paulson HL, Fishbeck KH. (1994) Trinucleotide expansion in neurological disease. *Ann. Neurol.* **36:** 814–822.

18. Kawaguchi Y, Okamoto T, Taniwaki M, et al. (1994) CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32. *Nature Genet.* **8:** 221–227.

19. Ambrose CM, Duyao MP, Barnes G, et al. (1995) Structure and expression of the Huntington's disease gene: Evidence against simple inactivation due to an expanded CAG repeat. *Somat. Cell Mol. Genet.* **20:** 27–38.

20. Li X-J, Li SH, Sharp AH, et al. A Huntington-associated protein enriched in brain with implications for pathology of Huntington's Disease. *Nature* (in press).