Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 6509–6513, July 1995
Biochemistry

Incorporation of glutamine repeats makes protein oligomerize: Implications for neurodegenerative diseases

KELVIN STOTT*, JONATHAN M. BLACKBURN*, P. J. G. BUTLER†, AND MAX PERUTZ*†

*Medical Research Council Centre for Protein Engineering and †Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England

Contributed by Max Perutz, April 11, 1995

ABSTRACT **Many transcription factors and some other proteins contain glutamine repeats; their abnormal expansion has been linked to several dominantly inherited neurodegenerative diseases. Having found that poly(L-glutamine) alone forms β-strands held together by hydrogen bonds between their amide groups, we surmised that glutamine repeats may form polar zippers, an unusual motif for protein–protein interactions. To test this hypothesis, we have engineered a Gly-Gln_{10}-Gly peptide into the inhibitory loop of truncated chymotrypsin inhibitor 2 (CI2), a small protein from barley seeds, by both insertion and replacement. Gel filtration resolved both mutant inhibitors into at least three fractions, which analytical ultracentrifugation identified as monomers, dimers, and trimers of the recombinant protein; the truncated wild-type CI2 formed only monomers. CD difference spectra of the dimers and trimers versus wild type indicated that their glutamine repeats formed β-pleated sheets, while those of the monomers versus wild type were more suggestive of type I β-turns. The CD spectra of all three fractions remained unchanged even after incubation at 70°C; neither the dimers nor the trimers dissociated at this temperature. We argue that the stability of all three fractions is due to the multiplicity of hydrogen bonds between extended strands of glutamine repeats in the oligomers or within a β-hairpin formed by the single glutamine repeat of each monomer. Pathological effects may arise when expanded glutamine repeats cause proteins to acquire excessively high affinities for each other or for other proteins with glutamine repeats.**

Five dominantly inherited neurodegenerative diseases have now been linked to abnormally expanded stretches of polyglutamine in the affected proteins: Huntington disease; spinal and bulbar muscular atrophy, also known as Kennedy disease; spinocerebellar ataxia type 1; dentatorubral–pallidoluysian atrophy (1, 2); and Machado–Joseph disease (3). All five diseases become more severe and begin earlier the longer the glutamine repeats. They are encoded by CAG trinucleotide repeats, which tend to lengthen in successive generations of affected individuals, especially in male transmission. After modeling had suggested that glutamine repeats might act as polar zippers, linking β-strands by hydrogen bonds between both their main-chain and side-chain amides, a synthetic peptide with the sequence Asp_2-Gln_{15}-Lys_2 was found to aggregate into pleated β-sheets (4–6). Encouraged by this result, we have now engineered glutamine repeats into the inhibitory loop of chymotrypsin inhibitor 2 (CI2) from barley seeds, a small monomeric protein of known structure (7), in order to determine whether they make this protein associate into oligomers.

We chose this small protein rather than any of those affected by the five neurodegenerative diseases, because none of those has yet been isolated in workable quantities. They are all very large, and it seemed doubtful that they could be expressed in soluble and native form by recombinant DNA technology. At this early stage, it seemed important above all to establish in principle whether incorporation of glutamine repeats makes proteins oligomerize, for which there has previously been no direct evidence from protein chemistry. Glutamine-rich domains have been found to be responsible for interaction between transcription factors, but it has not been proved that this is due solely to specific linkages between their glutamines.

MATERIALS AND METHODS

pJB2 was constructed by ligation of the 200-bp *Nde* I/*Hin*dIII fragment of pCI2 (8) with the 4.5-kb *Nde* I/*Hin*dIII fragment of pRH1090 (9). pCI2-Q10(ins), the vector used for expression of the loop insertion mutant of truncated CI2, was constructed by insertion of a synthetic DNA fragment encoding the Gly-Gln_{10}-Gly peptide into the *Nco* I restriction site at the Met-59 codon of the CI2 gene in pJB2; the insert was first generated by phosphorylating and annealing the two 39-mer oligonucleotides 5′-CATGGGTCAGCAGCAACAACAG-CAGCAACAACAGCAGGG-3′ and 5′-CATGCCCTGCT-GTTGTTGCTGCTGTTGTTGCTGCTGACC-3′ and was then ligated directly into a dephosphorylated, gel-purified partial *Nco* I digest of pJB2. pCI2-Q10(rep), the vector used for expression of the loop replacement mutant of truncated CI2, was constructed by substitution of the *Pst* I/*Hin*dIII fragment of the CI2 gene in pJB2 with an analogous fragment in which codons 54–61 were replaced with a stretch of DNA encoding the Gly-Gln_{10}-Gly peptide; the 140-bp substitute fragment was first generated by PCR, using pJB2 as template DNA and the two oligonucleotides 5′-AAAACTGCAGGA-CAAGCCAGAGGCGCAAATCATAGTTCTGCCGGTG-GGGCAGCAGCAACAACAGCAGCAACAACAGCAGG-GGCGGATCGACCGCGTCCG-3′ and 5′-GCCGCCAGGC-AAATTCTG-3′ as forward and reverse primers, respectively; after complete digestion with *Pst* I and *Hin*dIII, the purified fragment was ligated with the 4.6-kb *Pst* I/*Hin*dIII fragment of pJB2. For each of the three constructs, ligated DNA was electroporated into *Escherichia coli* strain NM554, cells were plated on rich medium containing chloramphenicol, colonies were screened by PCR, and one positive clone was restriction-mapped and sequenced to confirm its identity (10).

The soluble extracts, each from 1 liter of induced culture, were first precipitated with 70% saturated ammonium sulfate, dialyzed into 50 mM Tris·HCl (pH 8.6), passed through 1 ml of DE52 anion-exchange resin, and then concentrated to ≈5 ml over ultrafiltration membranes with a molecular weight cutoff of 3000. NaCl (150 mM) was added to the concentrated extracts before loading onto a HiLoad 26/60 Superdex 75 size-exclusion column for gel filtration by fast protein liquid chromatography. Protein was eluted with 50 mM Tris·HCl, pH 8.6/150 mM NaCl at 2.5 ml/min and monitored as absorbance at 280 nm.

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.

Abbreviation: CI2, chymotrypsin inhibitor 2.

RESULTS

CI2 consists of a single polypeptide chain of 83 residues, the first 18 of which are disordered in its crystal structure (Fig. 1); they are followed by a short β-strand, a three-turn α-helix, and another three β-strands associated into a pleated sheet, which forms a stable hydrophobic core with the α-helix. β-strands 2 and 3 are connected by a long and flexible solvent-exposed loop consisting of residues 54–62, including the inhibitory methionine (Met-59). We have made two mutants of a truncated form of CI2 from which the first 20 residues were deleted. In the first mutant, a sequence of 10 glutamine residues flanked by two glycines (for flexibility) was inserted into the loop at Met-59, while in the second mutant, residues 54–61 of the loop were replaced with this sequence. The parent vector pJB2 (Fig. 2) was constructed for expression of the truncated wild-type CI2 and was then used to construct vectors for expression of the two mutant inhibitors. Expression was induced in *E. coli* strain NM554 with 1 mM isopropyl β-D-thiogalactopyranoside, yielding ≈150 mg of recombinant protein per liter of culture after purification.

Fig. 3 shows the fractionation of soluble cell extracts of the truncated wild-type inhibitor and of the two mutants by gel filtration. Cell extract of the truncated wild-type inhibitor gave a broad heterogeneous fraction of large cellular proteins (fraction a), a clean concentrated fraction of the recombinant protein (fraction d), and a series of low molecular weight fractions (fraction e), which were later found to be nucleic acid, probably RNA. The cell extracts of the two mutant inhibitors, on the other hand, both yielded two additional fractions (fractions b and c) of higher apparent molecular weight than the main fraction of monomeric recombinant protein, but quite distinct from the broad heterogeneous fraction of large

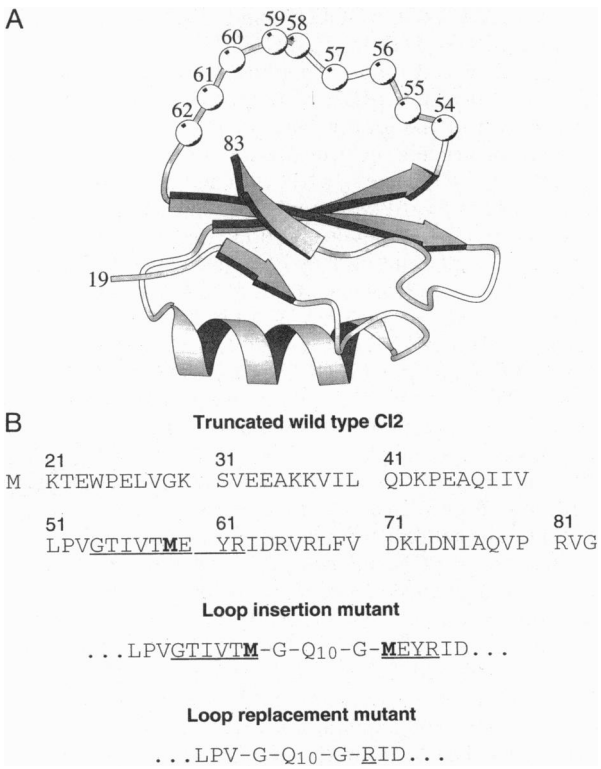

FIG. 1. (*A*) Structure of wild-type CI2; the first 18 residues are disordered and are not shown (7). Loop residues 54–62 are represented by spheres. The Gly-Gln$_{10}$-Gly peptide was inserted into the loop at Met-59 in one mutant, while the second mutant was generated by replacing residues 54–61 of the loop with this peptide. (*B*) Amino acid sequence of the truncated wild-type CI2 and of the two mutants. Original residues of the loop are underlined.

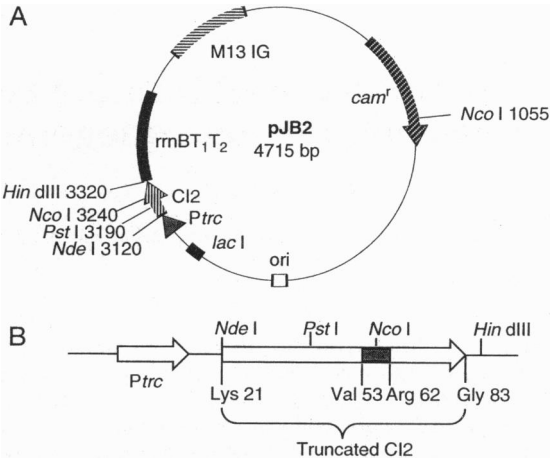

FIG. 2. (*A*) Plasmid map of the parent vector pJB2 used for expression of the truncated wild-type CI2. (*B*) Structure of the truncated CI2 gene in pJB2, including the relevant codons and restriction sites.

cellular proteins. SDS/PAGE of fractions b, c, and d of each mutant inhibitor produced a single clean band at the same position, close to that of the truncated wild-type inhibitor, showing that all three fractions consisted of the recombinant protein. This was confirmed by electrospray mass spectrometry, which gave essentially identical values for the molecular weights of each of the three fractions of the two mutant inhibitors, indicating that fractions b and c were higher-order aggregates of the recombinant proteins in fraction d. The observed molecular weights also verified their correct expression (Table 1).

The order of oligomerization in fractions b, c, and d of both mutant inhibitors was determined by analytical ultracentrifugation. Fig. 4 shows the results for the loop insertion mutant inhibitor, indicating apparent molecular weights for fractions b, c, and d of 26,000, 17,900, and 9900, respectively. These weights agree with those expected for trimer, dimer, and

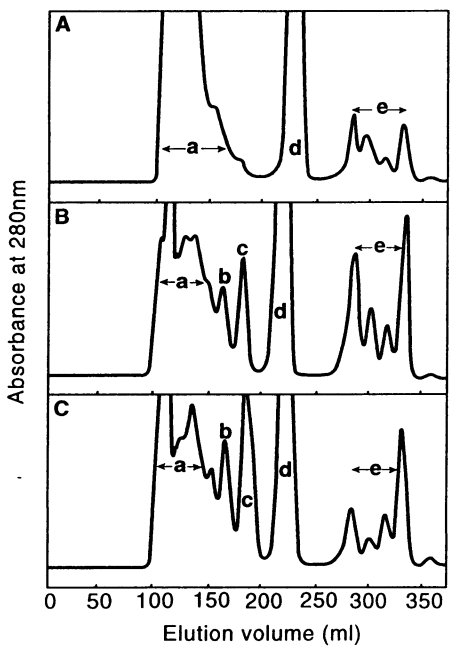

FIG. 3. Preparative gel filtration of soluble cell extracts from *E. coli* expressing the truncated wild-type CI2 (*A*), the loop insertion mutant (*B*), and the loop replacement mutant (*C*). Full absorbance scale of each elution profile is 2.0.

Table 1. Calculated vs. observed molecular weight

	Molecular weight	
	Calculated	Observed
Truncated wild-type CI2	7303.6	7304.1 (±0.7)
Loop insertion mutant	8830.3	8830.3 (±0.5)
Loop replacement mutant	7804.0	7804.0 (±0.3)

monomer (M_r, 26,500, 17,700, and 8,800). Similar results were obtained for the loop replacement mutant inhibitor (data not shown). Analysis of the fractions by fast protein liquid chromatography (Pharmacia) on a calibrated Superdex 75 HR 10/30 gel-filtration column produced peaks at elution volumes that were also compatible with the apparent molecular weights expected for trimer, dimer, and monomer (data not shown). Fractions c and d of each mutant inhibitor produced clean single peaks, whereas fraction b produced a peak with a small shoulder coinciding with the peak of fraction c, probably due to dimer as a contaminant from the less efficient isolation of the trimeric fraction. This is reflected by the decrease in apparent molecular weight of fraction b from trimer to dimer at low protein concentration in Fig. 4.

The CD spectra of all three fractions of either mutant were similar to the spectrum of the truncated wild-type inhibitor, as shown in Fig. 5 for the loop insertion mutant, which speaks against any major changes in tertiary structure. The difference spectrum between the trimer and the wild type shows a positive peak at 195 nm and a negative peak at 223 nm, coincident with the pleated β-sheet spectrum of the original oligopeptide Asp_2-Gln_{15}-Lys_2 (5). The difference spectrum of monomer versus wild type above 200 nm resembles the CD spectrum of type I β-turns, but it lacks its large positive peak at 195 nm (11). We have been able to reproduce it by heating a solution of the oligopeptide to 95°C and letting it cool slowly to 25°C (Fig. 5*C*). Heating may have caused the β-strands in the pleated sheets of the oligopeptide to dissociate and fold into hairpins; gel filtration revealed that these then reassociated on cooling. The similarity between the spectrum of the heated oligopeptide and the difference spectrum of the monomer versus wild type suggests that the glutamine repeat inserted in the CI2 may also fold into a hairpin. The difference spectrum of the dimer

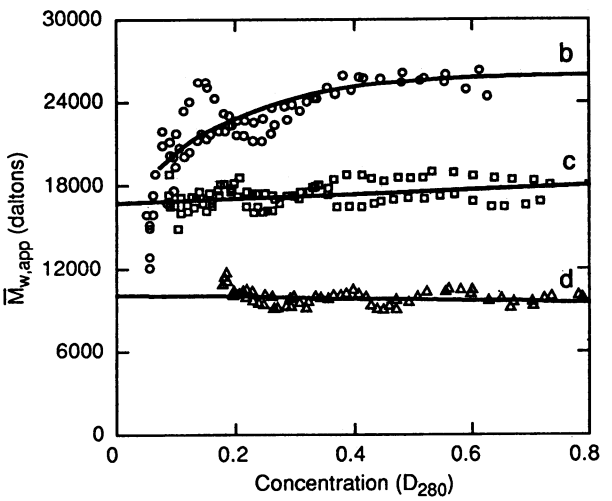

FIG. 4. Plots of apparent weight average molecular weights ($\bar{M}_{w,app}$) against concentration for the peaks b, c, and d from fast protein liquid chromatography fractionation. Molecular weights were determined by equilibrium centrifugation and analysis as described (12), assuming a partial specific volume of 0.73 ml/g. Curve b was fitted for nonideality, while curves c and d were drawn in arbitrarily to fit the overlaid data from separate experiments.

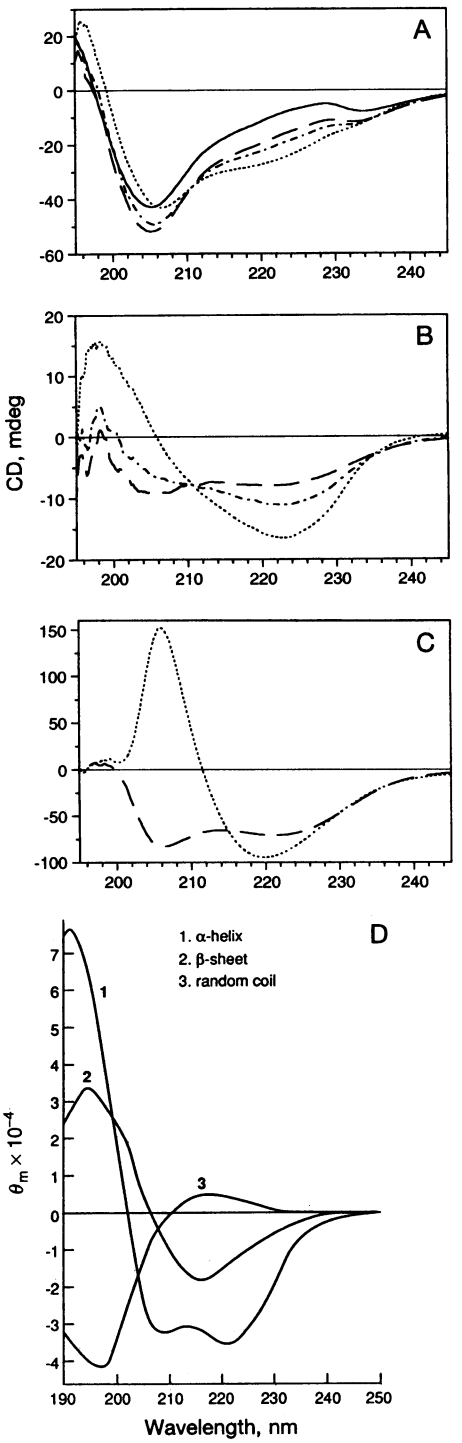

FIG. 5. (*A*) CD spectra of truncated wild-type CI2 (——) and of monomeric (— —), dimeric (- - -), and trimeric (- - -) fractions of the loop insertion mutant in 50 mM potassium phosphate buffer (pH 6.85) at 25°C. All three fractions and the wild type had an absorbance of 0.4 at 282 nm. With an extinction coefficient calculated from their amino acid composition of 6970 $M^{-1}\cdot cm^{-1}$, which was independent of the presence of the glutamines and of 6 M guanidinium chloride, this absorbance corresponded to a concentration of 57.4 μM. (*B*) Corresponding difference spectra of the three mutant fractions relative to the spectrum of the truncated wild-type CI2. (*C*) Spectrum of the oligopeptide Asp_2-Gln_{15}-Lys_2 before (- - -) and after (——) heating to 95°C. Spectra were recorded as the average of 30 individual spectra, each measured using a Jasco J-720 spectropolarimeter with a 0.1-cm-pathlength cell, at a scan rate of 20 nm/min. (*D*) Standard CD spectra of α-helix, β-sheet, and random coil structures of poly(L-lysine). (Reproduced with permission from ref. 13; copyright 1969, American Chemical Society.)

versus wild type is intermediate between the difference spectra of the trimer and the monomer versus wild type.

That the three different fractions of either mutant could be isolated suggests that they are all fairly stable; otherwise, a single fraction containing an equilibrium mixture of the three oligomerization states would have been produced. In fact, analytical gel filtration revealed that all three fractions of the loop insertion mutant remained pure even after storage at 4°C for 2 months. Those of the loop replacement mutant inhibitor were slightly less stable; over a period of weeks, a small fraction of the trimers of this mutant dissociated into monomers and dimers, and a small fraction of the dimers dissociated into monomers, while the monomers remained pure, even at a concentration of 2 mM. To compare the stability of the three fractions with that of the wild type, we followed their CD at 222 nm while raising the temperature from 2°C to 95°C (Fig. 6). The wild type was stable up to 80°C and denatured at 85°C, when an increase in absorption at 280 nm showed that it aggregated. The CD of the three fractions exhibited no transitions below 50°C, which showed that their glutamine zippers were stable to at least this temperature. The monomer showed only a single transition at 80°C, when it denatured and aggregated, as did the dimer and trimer, such that the CD curves of all three fractions converged at 90°C. The CD curves of the dimer and trimer indicate that they underwent structural transitions between 50°C and 70°C. To determine whether these corresponded to dissociation into monomers, we incubated all three fractions at 70°C for 1 hr and then cooled them slowly to 25°C, but instead of finding the CD spectra of the dimer and trimer converted to the spectrum of the monomer, they all remained unchanged, proving that their glutamine zippers remained intact even at 70°C. When we tried to establish a dynamic equilibrium between monomers and oligomers in various concentrations of urea, we found that the entire protein denatured cooperatively with its glutamine repeat at around 3.0 M urea.

The stability of all three fractions implies that both association and dissociation rates are extremely slow. The slow association rates of the monomers can be explained only by the presence of intramolecular hydrogen bonds within their glutamine repeats, which are likely to be bent into hairpins. The slow dissociation rates of the oligomers are clearly due to the multiplicity of intermolecular hydrogen bonds between extended β-strands of glutamine repeats. The conformation of β-strands would be the same in hairpins as in β-sheets, but hairpins in the monomers would be entropically favored over β-sheets of glutamine repeats in the oligomers. This may explain why monomers were the dominant species formed during overexpression in *E. coli*. In low expression, oligomers would be less likely to form, unless the glutamine repeats become long enough for the hairpins themselves to associate into β-sheets.

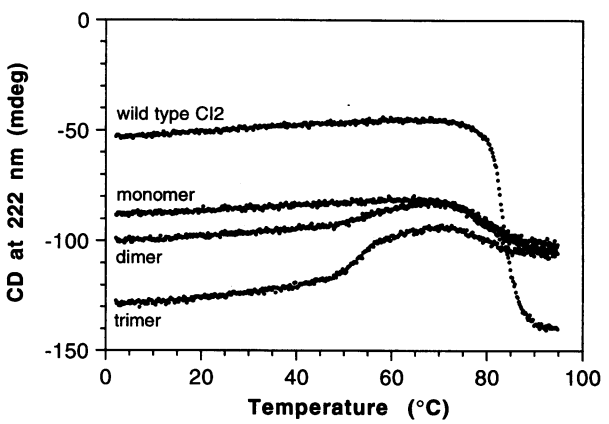

FIG. 6. Thermal denaturation of truncated wild-type CI2 and of the three fractions of the loop insertion mutant, each at a concentration of 28.7 μM in 50 mM potassium phosphate buffer at pH 6.85, followed by CD at 222 nm using a cell of 1-cm pathlength.

DISCUSSION

How common are glutamine repeats or glutamine-rich domains, where do they occur, and what is their function? On screening the Swiss-Prot data base for proteins with at least 20 glutamine repeats, Gerber *et al.* (14) found 33 of 40 top scoring proteins to be transcription factors. Tjian and his colleagues (15–19) have explored the role of glutamine-rich domains in the human transcription factor SP1. Transcription of reporter genes was enhanced by interaction between SP1s bound to GC-rich promoters 1.8 kb apart; the glutamine-rich domains proved essential for this interaction, which was vividly demonstrated in electron micrographs picturing the looping of the intervening DNA when two distantly bound SP1s were joined together. SP1 also interacted strongly with glutamine-rich TATA-binding-associated factors. Gerber *et al.* (14) measured activation of transcription by the Gal-4 chimeric protein, one part of which was linked to SP1 and the other to TATA-binding-associated factors or other proteins with either glutamine repeats or glutamine-rich domains. They all stimulated high levels of transcription. Stimulation increased with the number of glutamine repeats up to 40, remained constant up to 80 repeats, and then declined. These experiments suggest that glutamine repeats play a role in transcription similar to that of leucine zippers.

Only one of the proteins responsible for the five dominantly inherited neurodegenerative diseases is known to be a transcription factor. This is the androgen receptor responsible for Kennedy disease. The function of the Huntington disease protein is still unknown, because its amino acid sequence of >3100 residues (M_r, ≈350,000) shows no homology with other known proteins; contrary to earlier reports, it has now been located in the cytoplasm rather than the cell nucleus (20). C. A. Ross (Johns Hopkins Department of Psychiatry and Neuroscience) has informed us that a Western blot of monkey cortex fractionated on a nondenaturing gel and probed with anti-Huntington disease antibodies by G. Schilling and A. H. Sharp yielded a broad peak at M_r ≈700,000, showing that the protein either associated into dimers or associated with other proteins.

Several observations have proved that neural damage in diseases affected by expanded glutamine repeats is due to gain rather than loss of function. Our results show that glutamine repeats make proteins associate into stable oligomers. In transcription factors, expanded repeats may therefore lead to "aberrant transcriptional activity" (21), because these repeats acquire excessive affinities for each other or for complementary regulatory proteins with glutamine repeats. In cytoplasmic proteins, expanded glutamine repeats may lead to "wrong" interactions with other proteins. We hope that our proof of the function of glutamine repeats will serve as a step toward unraveling the molecular mechanism of these terribly distressing neurodegenerative diseases. Our discovery may also find uses in protein engineering; oligomers held together by glutamine repeats may serve as models for the design of new proteins.

We thank Prof. A. R. Fersht for suggesting the use of CI2 for this experiment, Fiona Suit for synthesizing our oligonucleotide primers, Dr. Ian Fearnley for determining the molecular weights of our constructs by electrospray mass spectrometry, Dr. Olga Perisic for helpful advice with their preparation, and Dr. Tony Johnson for determining the distribution of molecular weights of our oligopeptide before and after heating by gel filtration. K.S. is supported by a Medical Research Council Research Studentship, J.M.B. is supported by a Fitzwilliam College Research Fellowship, and M.P.'s research is supported by National Institutes of Health Grant HL31461.

1. Martin, J. B. (1993) *Science* **262,** 674–676.
2. Ross, C. A., McInnes, M. G., Margolis, R. L. & Li, S.-H. (1993) *Trends Neurosci.* **16,** 254–259.
3. Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., Akiguchi, I., Kimura, J., Narumiya, S. & Kakizuka, A. (1994) *Nat. Genet.* **8,** 221–228.
4. Perutz, M. P., Staden, R., Moens, L. & De Baere, I. (1993) *Curr. Biol.* **3,** 249–253.
5. Perutz, M. F., Johnson, T., Suzuki, M. & Finch, J. T. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 5355–5358.
6. Perutz, M. F. (1994) *Protein Sci.* **3,** 1629–1637.
7. McPhalen, C. A., Svendsen, I., Jonassen, I. & James, M. N. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7242–7246.
8. Jackson, S. E., Moracci, M., el Masri, N., Johnson, C. M. & Fersht, A. R. (1993) *Biochemistry* **32,** 11259–11269.
9. Baldwin, J. E., Blackburn, J. M., Heath, R. I. & Sutherland, J. D. (1992) *Bioorg. Med. Chem. Lett.* **2,** 663–668.
10. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
11. Perczel, A., Hollósi, M., Sündor, P. & Fasman, G. D. (1993) *Int. J. Pept. Protein Res.* **41,** 223–236.
12. Tennent, G. A., Butler, P. J. G., Hutton, T., Woolfitt, A. R., Harvey, D. J., Rademacher, T. W. & Pepys, M. B. (1993) *Eur. J. Biochem.* **214,** 91–97.
13. Greenfield, N. & Fasman, G. D. (1969) *Biochemistry* **8,** 4108–4116.
14. Gerber, H. P., Seipel, K., Georgiev, O., Höfferer, M., Hug, M., Rusconi, S. & Schaffner, W. (1994) *Science* **263,** 808–811.
15. Courey, A. J. & Tjian, R. (1988) *Cell* **55,** 887–898.
16. Courey, A. J., Holtzman, D. A., Jackson, S. P. & Tjian, R. (1989) *Cell* **59,** 827–836.
17. Pascal, E. & Tjian, R. (1991) *Genes Dev.* **5,** 1646–1656.
18. Su, W., Jackson, S., Tjian, R. & Echols, H. (1991) *Genes Dev.* **5,** 820–826.
19. Hoey, T. H., Weinzierl, R. O. J., Gill, G., Chen, J.-L., Dynlacht, B. D. & Tjian, R. (1993) *Cell* **72,** 247–260.
20. Sharp, A. H., Loev, S. J., Schilling, G., Steiner, J. P., Lo, A., Hedreen, J., Sisodia, S., Dawson, T. M., Snyder, S. H. & Ross, C. A. (1994) *Soc. Neurosci. Abstr.* **20,** 1648.
21. Mhatre, A. N., Trifiro, M. A., Kaufman, M., Kazemi-Esfarani, P., Figlewicz, D., Rouleau, G. & Pinsky, L. (1993) *Nat. Genet.* **5,** 184–188.