A Huntington's disease (HD) phenocopy leads to the identification of the HD toxin as polyalanine. The relationship between repeat length and age of onset can be reframed as a rate equation and the toxic polyalanine length is found to be about 30.6.

Eugen Tarnow

Avalon Business Systems, Inc., 18-11 Radburn Road, Fair Lawn, NJ 07410, USA,

etarnow@avabiz.com

Abstract

Huntington's disease (HD) is one of the most well defined "repeat diseases", associated with a short repeated genetic sequence, CAG.

First, taking into account that a phenocopy of HD has a different repeat that is associated with a different gene, I suggest that the gene is not important for HD, only the repeat sequence is important, in agreement with Lee et al (2019) who reached the same conclusion using a GWAS technique.

Second, taking into account that a phenocopy of HD has a CTG repeat rather than a CAG repeat, and that the toxin should be the same for both disease types and that the third base in a codon is the least important, I suggest that the reading frame is shifted for the repeat expansions and that the A/T substitution takes place on the third base. The most likely sense and antisense reading frames are then $(GCA)_n$ and $(GCT)_n$ and $(GCT)_n$ and $(GCA)_n$ and the corresponding amino acid is polyalanine.

Third, the more repeats, the earlier the HD onset (Brinkman et al, 1997; Wexler, 2004). I suggest that this relationship can be thought of as a rate equation. If the concentration is proportional to the probability of creating a polyalanine of length m in a repeat expansion of length n, the corresponding equation is borne out by the data on age of onset and repeat length and m is found to be about 30.6. This explains for the first time, at least approximately, why HD is not active unless there are at least 36 CAG repeats.

If true, HD may be the first disease where frameshifting is the cause of the disease.

Keywords: Huntington's disease, CAG, kinetics, repeat expansions.

Introduction

Because neurons are not regenerated, neurodegenerative diseases are different from other diseases in that an extremely slow and unlikely process can, after a lifetime, become deadly.

Huntington's disease (HD) is one of more than 20 genetic repeat diseases (La Spada & Taylor, 2010). The repeated sequence is CAG attached to the HTT gene. There is also a phenocopy (HDL2) with a repeat sequence CTG attached to the JPH3 gene as well as other genetic causes (Mariani et al, 2016).

It is thought that the repeat sequence "results in production of HTT protein with an expanded polyglutamine tract (polyQ), leading to pathogenic HTT protein conformers that are resistant to protein turnover, culminating in cellular toxicity and neurodegeneration" (Dickey & La Spada, 2018; however, this is controversial, see Ratovitski et al, 2016; Veitch et al, 2007). Whether this explains the CTG phenocopy is not obvious since they are located next to two different genes (HTT and JPH3) and CTG codes for leucine, not glutamine and glutamine is polar and leucine is non-polar.

A second problem with the theory of "pathogenic HTT protein conformers" is as follows. If a protein misfolds, it would seem reasonable that there should be a critical number of repeats that causes a misfold and more repeats beyond this critical number should have little effect. In HD, however, that the age of onset is a monotonic function of the number of repeats and there is no evidence of a critical number of repeats other than that the age of onset increases: extrapolation suggests that at 37 repeats the median age of onset should be 79 and at 35 it should be 95 (Fig. 1, data from Brinkman et al, 1997).
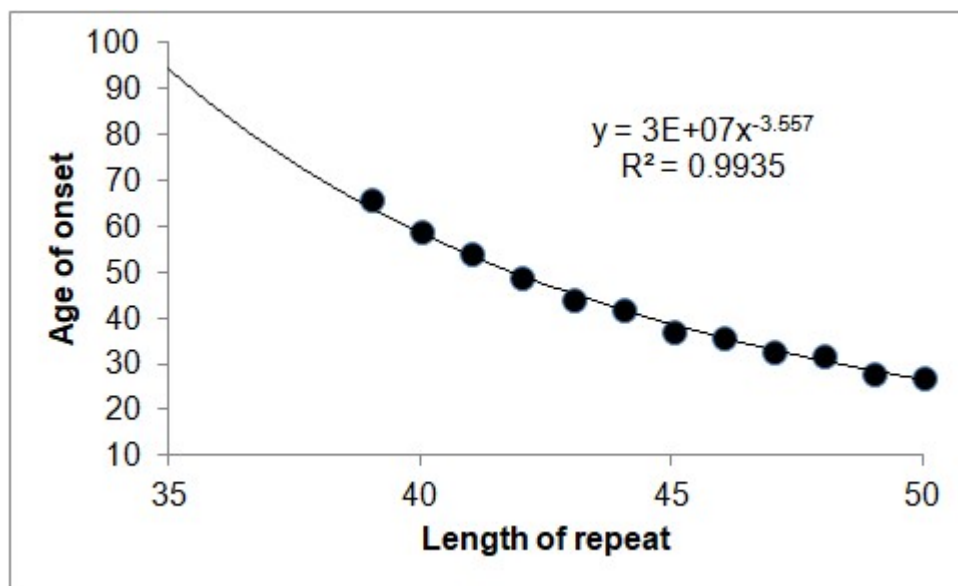
Figure 1. HD median age of onset (50% of individuals affected) as a function of the repeat length. Data from Brinkman et al (1997), fitted by the current author. The line on the left is an extrapolation of the data.

In addition to the difficulty in identifying the toxin in HD, there is some controversy as to how long the repeat has to be in order to cause HD. Banez-Coronel et al (2015) write that "expansions >40 repeats are fully penetrant and longer alleles are associated with earlier, sometimes juvenile onset and increased disease severity". Brinkman et al (1997) find that expansions over 41 are fully penetrant and that only a portion of those with 36-41 repeats showed symptoms within a life time.

In this paper I will propose three new ideas to assist in the search for the HD toxin. Together they identify the toxin to be polyalanine repeated about 30.6 times.

Sample & Method

The data used is from Brinkman et al (1997) who describes it:

"For the purpose of this study we used those individuals with CAG expansions of the upper allele that were >28, comprising 728 affected and 321 asymptomatic at-risk individuals from 473 families, whose age at onset or oldest age while still asymptomatic could be ascertained."

"An accurate assessment of the age at onset was performed through both a retrospective review of patient charts and telephone interviews with patients, family members, genetic counselors, and physicians. Age at onset was defined as the first time at which a patient had either neurological or psychiatric symptoms that represented a permanent change from the normal state. The age used for analysis of all asymptomatic individuals was the oldest age when his or her clinical status was last directly confirmed, either at the genetics clinic in Vancouver or by the local, attending physician. Particular attention was paid to confirmation of current age and clinical status of all asymptomatic, at-risk individuals in the HD database who were >65 years of age."

Microsoft Office Excel 2007 was used to curve fit the median age as a function of the number of CAG repeats.

Results: It is the CAG expansion, not the gene.

If we assume that HD and its phenotype copy has the same pathology, it means that it is the expansion, not the different neighboring genes HTT and JPH3, that is responsible for the pathology. This is particularly true if we insist that the two expansions are the same as I do in the next section.

Results: Rule for Reading Frame Assignment

I propose that for repeat diseases with phenocopies that have shifted repeats, reading frame errors are crucial and that the relevant reading frames are those that cause both repeats to code for the most alike protein. For a three-base codon repeat sequence in which the phenocopy has a single base substituted, the relevant reading frame is the one in which the substitution occurs on the third codon base, the least important one.

Thus for the CAG and CTG expansions of HD and its phenocopy, the reading frame is changed so that the codon base substitution occurs in the third base instead of the second: instead of $(CAG)_n$ and $(CTG)_n$ it is $(GCA)_n$ and $(GCT)_n$. Both GCA and GCT code for the same amino acid – alanine.

While a change in reading frame is unlikely, I next suggest that the correlation of repeats with HD age of onset may be the kinetic equation of the corresponding biochemical reaction creating the toxin. Such a slow reaction arising from an unlikely reading frame error may be at the heart of HD.

Results: A Clue from Reaction Kinetics

I suggest that the relationship between repeat length and age of onset can be reframed in terms of chemical kinetics which can then give us information about chemical reaction taking place. I calculate the speed of the reaction assuming that the onset of the disease happens at a fixed critical concentration:

speed of reaction*age=critical concentration of toxin (equation 1)

Next I need to convert the length of the repeat into a concentration for the equation to be a kinetic equation. One way is to assume that the rate limiting step is the reading frame error resulting in a poly alanine molecule of length m. There are n-1-m positions in which this reading frame error can occur (because of the shift there are at most n-1 positions and in order to fit m alanines the number of positions is limited to n-1-m), the probability of creating this molecule is proportional to n-1 -m:

(n-1-m)*age=const

or

n=1+m+const/age (equation 2)

If I replot Fig. 1 in Fig. 2 with the y axis as 1/age the curve fulfills equation 2 and I identify m=30.6:
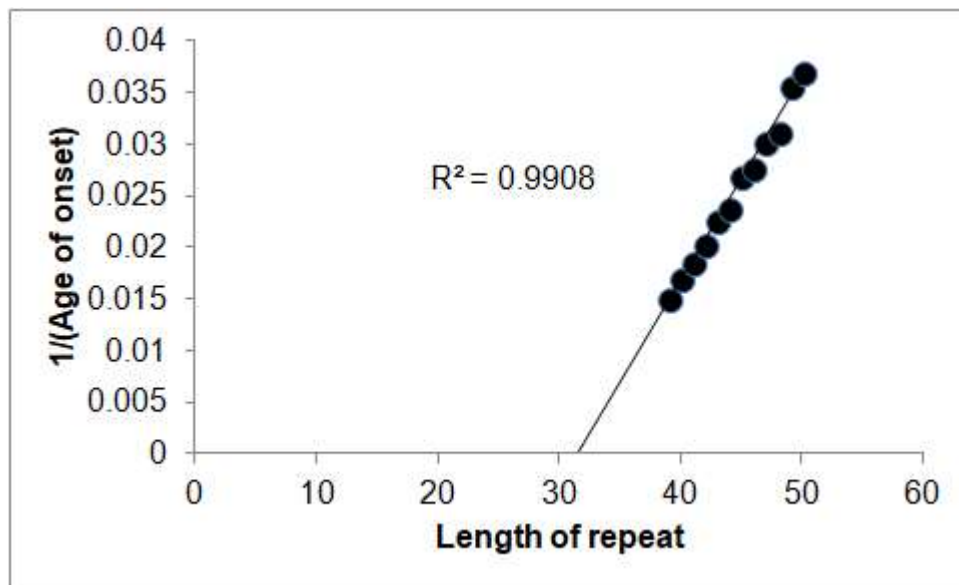
Fig. 2. The inverse of the (age of onset) versus the length of the repeat. The intersection with the x-axis identifies the length of the poly-alanine to be about 30.6 alanine units.

I calculate the age of onset as a function of n and find the following formula:

Age of onset = 505/(n-31.6) (equation 3)

This is an alternative to the exponential fit proposed by Kaplan et al but, in the mind of this author, not borne out by their theory (2007). The two fits are compared in Fig. 3.
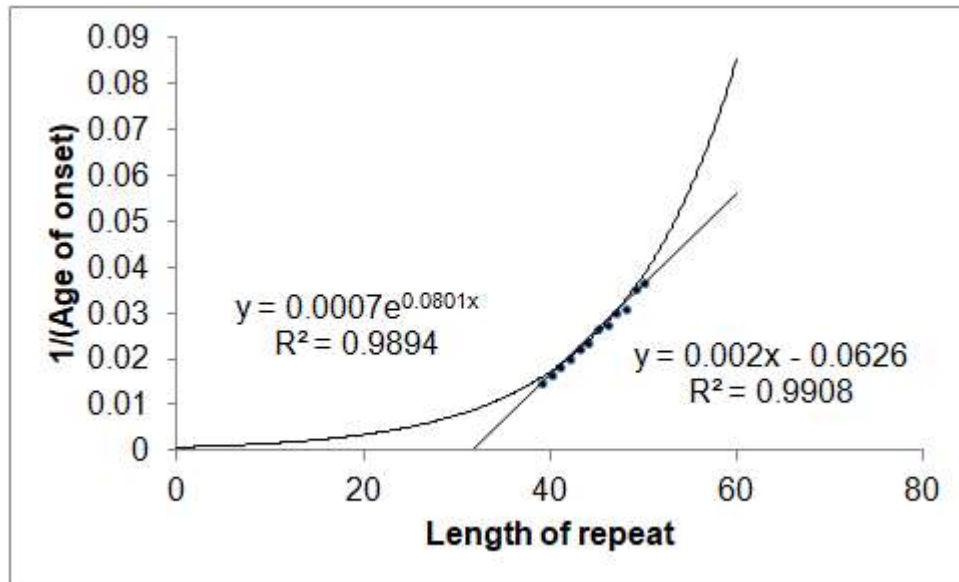
Fig. 3. Exponential fit and the fit from equation 3 compared. To distinguish the two fits, and perhaps settle the kinetics, it may be necessary to probe the extremes of repeats at less than 39 and more than 50.

Discussion

I proposed in three steps that the toxin in HD is polyalanin repeated about 30.6 times. First I suggested that that because there is a phenocopy of HD in which the repeat expansion occurs at a different gene (JPH3 instead of HTT), that the gene is not important, only the repeat expansion is. Second, because of the repeat expansions of HD and its phenocopy are somewhat different, the disease causing agent has to be the same which limits it to frameshifted polyalanines. Third, I interpreted the relationship between the repeat length and the age of onset as a kinetic equation for the creation of the polyalanine. I converted the repeat length into a concentration by noting that the probability of creating a frameshifted polyalanine of length m in a repeat expansion of length n is proportional to n-1-m. I showed that the corresponding linear equation between the inverse of the age of onset and the length of the repeat n is borne out by the data and allows me to identify the repeat length of the HD toxic polyalanine to be about 30.6.

There is previous literature that also suggests that the Huntingtin gene is not important: Lee et al (2019) arrived at this conclusion in a completely different way and write that "The effects of variable, glutamine-encoding, CAA interruptions indicate that a property of the uninterrupted HTT CAG repeat sequence, distinct from huntingtin's polyglutamine segment, dictates the rate at which HD develops".

Differently, "The number of consecutive glutamines in huntingtin is not equal to the number of consecutive CAG repeats because CAA interruptions also code for glutamine. Consequently, two individuals with the same CAG repeat length can produce different numbers of consecutive glutamines in their huntingtin protein, depending on the presence or absenceof CAA codons. The length of the uninterrupted CAG repeat turns out to be a better predictor of age-at-onset than is the length of the polyglutamine tract, indicating that the driver of the rate at which HD develops is a property of the uninterrupted CAG repeat rather than the length of encoded polyglutamine." (James Gusella, one of the authors of the Lee et al, 2019, paper, private communication). This is also consistent with my point that the toxin is not polyglutamine but polyalanine. Note that the frameshifted CAA would be ACA which codes for threonine, not alanine, and would consequently break up the polyalanine.

There is previous literature showing that HD includes frameshifted products including polyalanine (Toulouse et al 2005;Davies & Rubinsztein, 2006; Banez-Coronel et al, 2015) and the toxicity of polyalanines has been reviewed elsewhere (Amiel et al, 2004).

A polyalanine molecule of about 30.6 units provides the first explanation as to why at least 36 repeats are needed for HD to penetrate: several fewer repeats than 36 would make the creation of a 30 or 31 unit polyalanine molecule impossible with the reaction proposed and since the reaction creating the toxin is proportional to (n-1-m), the closer we get to m=n the much slower the reaction proceeds.

Graveland et al (1984) write that "the morphology of dendrites of medium-sized spiny neurons was markedly altered by the appearance of recurved endings and appendages, a decrease or increase in the density of spines, and abnormalities in the size and shape of spines". Is it possible that this is caused by a polyalanine repeat?

In at least one other neurodegenerative disease the most glaring pathogen is not the toxin: In Alzheimer's disease, amyloid plaques, the most obvious pathology thought to be toxic to neurons and thought to appear in late life coinciding with the dementia, were removed from patients but did not result in any clinical improvement (Drachman, 2014; Giacobini & Gold, 2013). A new suggestion is that Alzheimer's disease begins already in childhood or in puberty not with amyloid plaques but with abnormal tau protein deposits in an extremely slow process (Braak & Del Tredici, 2012). It may be that HD is similarly not caused by the HTT protein with an expanded polyglutamine tract but rather by a previously unlikely candidate resulting from an extremely slow and unlikely reading frame shift.

If the HD toxin is indeed caused by frame shifting, it is the first identified disease in which frame shifting is the underlying cause. Other diseases in which frame shifting is a factor including some cancers, Chron's disease, cystic fibrosis, HIV, etc.

Problems with this theory include:

1. It does not explain why homozygotes have the same age of onset as heterozygotes, rather my theory suggests that homozygotes should have half the age of onset for the same repeat length N.
2. It does not explain why other CAG repeat diseases have different symptoms.

Bibliography

Amiel, J., Trochet, D., Clément-Ziza, M., Munnich, A., & Lyonnet, S. (2004). Polyalanine expansions in human. Human molecular genetics, 13(suppl_2), R235-R243.

Bañez-Coronel, M., Ayhan, F., Tarabochia, A. D., Zu, T., Perez, B. A., Tusi, S. K., ... & Yachnis, A. T. (2015). RAN translation in Huntington disease. Neuron, 88(4), 667-677.

Braak, H., & Del Tredici, K. (2012). Where, when, and in what form does sporadic Alzheimer's disease begin? Current opinion in neurology, 25(6), 708-714.

Brinkman, R. R., Mezei, M. M., Theilmann, J., Almqvist, E., & Hayden, M. R. (1997). The likelihood of being affected with Huntington disease by a particular age, for a specific CAG size. American journal of human genetics, 60(5), 1202.

Deckel, A. Wallace, Paige Volmer, Ronald Weiner, Keith A. Gary, Jonathan Covault, Daniel Sasso, Nurit Schmerler, Daniela Watts, Zhige Yan, and Ilana Abeles. "Dietary alanine alters time of symptom onset in Huntington's disease transgenic mice." Brain research 875, no. 1-2 (2000): 187-195.

Dickey, A. S., & La Spada, A. R. (2018). Therapy development in Huntington disease: from current strategies to emerging opportunities. American Journal of Medical Genetics Part A, 176(4), 842-861.

Drachman, D. A. (2014). The amyloid hypothesis, time to move on: Amyloid is the downstream result, not cause, of Alzheimer's disease. Alzheimer's & Dementia, 10(3), 372-380.

Graveland, G. A., Williams, R. S., & DiFiglia, M. (1985). Evidence for degenerative and regenerative changes in neostriatal spiny neurons in Huntington's disease. Science, 227(4688), 770-773.

Giacobini, E., & Gold, G. (2013). Alzheimer disease therapy—moving from amyloid-β to tau. Nature Reviews Neurology, 9(12), 677.

Kaplan, S., Itzkovitz, S., & Shapiro, E. (2007). A universal mechanism ties genotype to phenotype in trinucleotide diseases. PLoS computational biology, 3(11), e235.

La Spada, A. R., & Taylor, J. P. (2010). Repeat expansion disease: progress and puzzles in disease pathogenesis. Nature Reviews Genetics, 11(4), 247.

Lee, J. M., Correia, K., Loupe, J., Kim, K. H., Barker, D., Hong, E. P., ... & Pinto, R. M. (2019). Huntington's disease onset is determined by length of uninterrupted CAG, not encoded polyglutamine, and is modified by DNA maintenance mechanisms. bioRxiv, 529768.

LEOPOLD, N. A., & PODOLSKY, S. (1975). Exaggerated growth hormone response to alanine infusion in Huntington's disease. The Journal of Clinical Endocrinology & Metabolism, 41(1), 160-163.

Mariani, L. L., Tesson, C., Charles, P., Cazeneuve, C., Hahn, V., Youssov, K., ... & Peuvion, J. N. (2016). Expanding the spectrum of genes involved in Huntington disease using a combined clinical and genetic approach. JAMA neurology, 73(9), 1105-1114.

Podolsky, S., & Leopold, N. A. (1977). Abnormal glucose tolerance and alanine tolerance tests in Huntington's disease. Gerontology, 23(1), 55-63.

Ratovitski, T., Chaerkady, R., Kammers, K., Stewart, J. C., Zavala, A., Pletnikova, O., ... & Ross, C. A. (2016). Quantitative proteomic analysis reveals similarities between Huntington's disease (HD) and Huntington's disease-like 2 (HDL2) human brains. Journal of proteome research, 15(9), 3266-3283.

Toulouse, A., Au-Yeung, F., Gaspar, C., Roussel, J., Dion, P., & Rouleau, G. A. (2005). Ribosomal frameshifting on MJD-1 transcripts with long CAG tracts. Human molecular genetics, 14(18), 2649-2660.

Veitch, N. J., Ennis, M., McAbney, J. P., The, U. S., Shelbourne, P. F., & Monckton, D. G. (2007). Inherited CAG· CTG allele length is a major modifier of somatic mutation length variability in Huntington disease. dna repair, 6(6), 789-796.

Wexler, N. S. (2004). Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. Proceedings of the National Academy of Sciences, 101(10), 3498-3503.

Wojciechowska, M., Olejniczak, M., Galka-Marciniak, P., Jazurek, M., & Krzyzosiak, W. J. (2014). RAN translation and frameshifting as translational challenges at simple repeats of human neurodegenerative disorders. Nucleic acids research, 42(19), 11849-11864.