Article

The New Face of Dynamic Mutation—The CAA [CAG]n CAA CAG Motif as a Mutable Unit in the *TBP* Gene Causative for Spino-Cerebellar Ataxia Type 17

Dorota Hoffman-Zacharska [1,2,*] and Anna Sulek [2,3]

1   Department of Medical Genetics, Institute of Mother and Child, 02-106 Warsaw, Poland
2   Institute of Psychiatry and Neurology, 02-957 Warsaw, Poland; anna.sulek@lazarski.pl
3   Faculty of Medicine, Lazarski University, 02-662 Warsaw, Poland
*   Correspondence: dorota.hoffman@imid.med.pl; Tel.: +48-22-3277313

Abstract: Since 1991, several genetic disorders caused by unstable trinucleotide repeats (TNRs) have been identified, collectively referred to as triplet repeat diseases (TREDs). They share a common mutation mechanism: the expansion of repeats (dynamic mutations) due to the propensity of repeated sequences to form unusual DNA structures during replication. TREDs are characterized as neurodegenerative diseases or complex syndromes with significant neurological components. Spinocerebellar ataxia type 17 (SCA17) falls into the former category and is caused by the expansion of mixed CAA/CAG repeats in the *TBP* gene. To date, a five-unit organization of this region [(CAG)3 (CAA)$_3$] [(CAG)n] [CAA CAG CAA] [(CAG)n] [CAA CAG], with expansion in the second [(CAG)n] unit being the most common, has been proposed. In this study, we propose an alternative organization scheme for the repeats. A search of the PubMed database was conducted to identify articles reporting both the number and composition of GAC/CAA repeats in *TBP* alleles. Nineteen reports were selected. The sequences of all identified CAG/CAA repeats in the TBP locus, including 67 cases (probands and b relatives), were analyzed in terms of their repetition structure and stability in inheritance, if possible. Based on the analysis of three units [(CAG)$_3$ (CAA)$_2$] [CAA (CAG)n CAA CAG] [CAA (CAG)n CAA CAG], the organization of repeats is proposed. Detailed analysis of the CAG/CAA repeat structure, not just the number of repeats, in *TBP*-expanded alleles should be performed, as it may have a prognostic value in the prediction of stability/instability during transmission and the possible anticipation of the disease.

Keywords: hereditary spinocerebellar ataxia; SCA17; *TBP* gene; TATA-binding protein; dynamic mutation; CAA/CAG repeat expansion

1. Introduction

The human genome contains numerous repeated sequences varying in complexity and organization, ranging from tandemly repeated satellites, typically localized in specific chromosomal regions, to interspersed repeats that generally represent inactive transposable elements but also active RNA genes. The simplest and most common form of these sequences is microsatellites, also known as simple/short tandem repeats (STRs) or simple sequence repeats (SSRs). Microsatellites consist of tandemly repeated DNA units with a repeat length of up to six nucleotides. They are highly polymorphic in populations mainly due to their length variability, which reflects changes in the number of repeated units at a *locus*.

Studies have shown that the expansion of repeat sequences of various lengths can be associated not only with human cancers but also with a growing number of hereditary diseases. Since 1991, several genetic disorders caused by this mechanism, known as *dynamic mutation*, have been described [1]. These disorders, referred to as a family of triplet repeat expansion diseases (TREDs), are characterized as neurodegenerative diseases or complex

syndromes with a significant neurological component. The first group is caused by the expansion of [CTG]n or [CAG]n repeats in the open reading frame (ORF), and the second by the expansion of [CGG]n or [CCG]n in 5′ or 3′ untranslated regions (UTRs) or [GAA]n or [TTC]n in introns of the respective genes [1,2]. Among the ten possible trinucleotide combinations, only these three have been demonstrated to undergo dynamic mutation, resulting in genetic diseases. Their combinations, as well as a less common one, [GAC]n or [CAG]n, can form stable secondary structures in vitro if they are long enough [3–5]. These sequences can adopt several unusual DNA structures, such as hairpins, triplexes, quadruplexes, and slipped structures, whose formation may disrupt DNA metabolism and serve as critical intermediates in dynamic mutation events [4,6]. The importance of such structure formation for repeat tract instability has been confirmed in vivo in *E. coli* and *S. cerevisiae* models [5]. TNRs are characterized by instability during transmission to the next generation as well as in somatic cells, with various DNA processes, including replication, recombination, repair, and transcription, influencing their stability. The precise mechanisms by which these systems interact to produce expansion or contraction remain unresolved, but the basic mechanism accepted for TNRs' instability is slipped strand mispairing during replication [6]. In this model, the formation of slipped DNA structures is proposed, leading to repeat expansion or deletion depending on strand localization related to the orientation effect of the sequence. However, this simple model does not explain the characteristic features of human TNR instability, such as the polarity of expansion (usually at the 3′ end of the track) and mutational bias towards expansion rather than contraction [3]. Another model for TNR expansion proposes the formation of FEN-1-resistant secondary structures at the 5′ flap end of the Okazaki fragment. The single-stranded DNA 5′-flap end of Okazaki fragments, originating from strand displacement during lagging strand synthesis, can be removed by flap endonuclease 1 (FEN-1). However, if these ends form structures like foldbacks or hairpins, this process is inhibited. It is also inhibited by the complex of the DNA-flap/single-strand binding protein (SSBP), which could be involved in the unfolding of unusual DNA structures [7]. Experimental systems show that mutations in FEN-1 and its *S. cerevisiae* ortholog RAD27 gene enhance microsatellite instability and spontaneous mutations (mainly sequence duplications). Such mutants have increased recombination rates and require a functional double-stranded break (DSB) repair system to survive [8,9]. These data suggest the additional involvement of recombination processes and replication perturbations in TNR instability. Recent studies indicate that expansion could occur by multiple processes and at different stages of germ-line development: in the pre-meiotic stage by replication polymerase slippage and DNA repair during meiosis by DSB repair and post-meiotically by DNA damage repair [10]. Generally, two types of factors may influence trinucleotide instability: trans-acting factors, including those involved in DNA replication and repair (FEN1, Msh2, Msh3, Msh6), and other factors reflecting specific properties of the loci, such as the presence of CpG islands, orientation, proximity to the replication origin, and the number and configuration of the repeats themselves (pure TNRs are more prone to instability during replication) [11].

The aim of this paper is to conduct a comparative analysis of the repeat region of the *TBP* gene (OMIM: 600075), where a dynamic mutation causes spinocerebellar ataxia type 17 (SCA17; OMIM: 607136). SCA17 is classified as a neurodegenerative TRED and is caused by the expansion of mixed CAA/CAG repeats in the gene-encoding TATA-binding protein (TBP). TBP is a general transcription factor and a component of TFIID, which is a transcription complex that regulates the expression of most eukaryotic genes transcribed by RNA polymerase II, as well as polymerase I and III transcription complexes (SL1 and TFIIIB, respectively) [12]. Neurodegenerative TREDs are caused by the expansion of [CAG] repeats in the open reading frames (ORFs) of the respective genes, leading to long polyglutamine (polyQ) tracts in the proteins. These disorders are thus known as polyglutamine (polyQ) diseases. The [CAG]n stretches in TRED genes are typically pure, but there are exceptions. For instance, non-pathogenic alleles of the spinocerebellar ataxia-related genes *ATXN1* (SCA1; OMIM 164400) and *ATXN2* (SCA2; OMIM 183090) have tracts interrupted by

1–3 CAT (His) codons in *ATXN1* and CAA (Gln) codons in *ATXN2* [13–15]. The absence of such interruptions in expanded pathogenic variants suggests their role in maintaining regional stability. A similar role is postulated for the penultimate CAA (Gln/Q) interruption at the 3′ end of the [CAG]n tract in the *HTT* gene (HD; OMIM 143100). Even when present in most expanded alleles, this codon can contribute to instability and expansion into the pathogenic range if mutated in alleles of an intermediate [CAG]n length [16].

The polyQ-coding region in the *TBP* (SCA17) gene is more complex, comprising both CAG and CAA Gln codons in normal and expanded alleles. Since Koide et al. described a sporadic case of cerebellar ataxia with pyramidal signs and severe intellectual impairment associated with CAG/CAA expansion in the *TBP* gene in 1999 [17], many SCA17 cases have been reported, some with detailed repeat sequence analysis. A large population study by Gostout et al. in 1993 [18] identified non-pathogenic alleles corresponding to 25–42 glutamine residues, finding 20 different alleles, most encoding 32–39 glutamines. Based on the sequence of 157 independent alleles, a five-unit organization of the *TBP* polyQ coding region, $[(CAG)_3 (CAA)_3] [(CAG)nI] [CAA CAG CAA] [(CAG)nII] [CAA CAG]$, with polymorphic blocks of pure (CAG)n repeats as the primary sites of repeat number variation were proposed [18]. Later studies established the SCA17 polyQ pathogenic range as 41 or more repeats, with reduced penetrance in the 41–48 repeat range, which is especially controversial in the 41–44 repeat range, where pathogenic variants in the *STUB1* gene coexist. This novel mechanism was defined as digenic *TBP/STUB1*-related SCA17 [19]. Recently, however, this mechanism has been questioned, and it was shown that intermediate $TBP_{40–49}$ alleles act as disease modifiers of SCA48 (OMIM 618093) caused by *STUB1* mutations rather than a STUB1/TBP digenic model [20]. The fully pathogenic range is set at 49 repeats or greater, with the largest known allele containing 66 repeats [17,18]. In some SCA17 cases, both the repeat number and the sequence of the mutated region have been analyzed [17,18,21–39]. The available data suggest three possible *TBP* gene mutations: the expansion of the CAG repeats in (CAG)nII, partial deletion of the region resulting in only one unstable (CAG)n tract, and partial duplication of the repeats region resulting in an expanded polyQ domain. Among the published cases, the most common form of mutation is the expansion in the [(CAG)nII] unit. Based on the available data and identified SCA17 cases, we propose a different view of the region, which is considered a three-unit region, $[(CAG)_3 (CAA)_2] [CAA (CAG)n CAA CAG] [CAA (CAG)n CAA CAG]$, with different susceptibility to expansions.

2. Results

Based on the PubMed search, which was performed to identify reports presenting CAG/CAA repeat region sequences of the *TBP* gene, 19 publications were chosen (listed in Table 1). The analysis of CAG/CAA repeat organization schemes of the SCA17/*TBP* gene was performed for published data as well as for probands and their relatives identified among HD/SCA patient cohorts of Polish origin diagnosed in the Dept. of Genetics Institute of Psychiatry and Neurology [38], unpublished data, Figure 1.

Sequences of all identified CAG/CAA repeats in the *TBP* locus—78 cases—(probands and relatives) were collected and are presented in Table 1. Five different types of repeat configurations were identified, which occurred with different frequencies in the study cohort and were characterized by different inheritance stability. The most common and stable in transmission was configuration $(CAG)_3 (CAA)_3 (CAG)_9$ CAA CAG CAA $(CAG)_n$ CAA CAG—65.4% cases, and the second $(CAG)_3 (CAA)_3 (CAG)_n$ CAA CAG—23%. This configuration showed instability in transmission from 1 to 7 (CAG)n repeats. Three other configurations, much rarer, were also identified $(CAG)_3 (CAA)_3 (CAG)_9$ CAA CAG CAA $(CAG)_n$ CAA CAG CAA $(CAG)_n$ CAA CAG—6.4%, $(CAG)_3 (CAA)_3 (CAG)_{12}$ CAA CAA $(CAG)_{13}$ CAA $(CAG)_{16}$ CAA CAG—2.6% and $(CAG)_3 (CAA)_3 (CAG)_9$ CAA CAG CAA $(CAG)_9 (CAA)_3 (CAG)_9$ CAA CAG CAA $(CAG)_n$ CAA CAG, with 2.6% identified de novo in paternal transmission. Based on the analysis of all those sequences, we propose an alternative model of *TBP* CAG/CAA organization and expansion presented in Table 2.

Table 1. Characterization of the CAA/CAG repeats range and configuration in normal and expanded TBP alleles (previously reported). In the case of the expanded alleles, the number of analyzed cases and transmission is pointed out, as well as the stability during transmission and the possible mechanism of mutation.

CAG/CAA Repeat Structure in TBP Gene	Repeats No.	Analyzed Cases	Stability	Possible Mutation Mechanism	Reference
normal					
$(CAG)_3 (CAA)_3 (CAG)_{nI} CAA CAG CAA (CAG)_{nII} CAA CAG$	29–40	157 unrelated alleles		$(CAG)_{nI/nII}$ polymorphism	[18]
expanded					
$(CAG)_3 (CAA)_3 (CAG)_9 CAA CAG CAA (CAG)_9 (CAA)_3 (CAG)_9 CAA CAG CAA (CAG)_{19} CAA CAG$	63	single case (de novo paternal transmission)	-	partial duplication	[17]
$(CAG)_3 (CAA)_3 (CAG)_9 CAA CAG CAA (CAG)_{26} CAA CAG$	46	3 gen. (6 family members)	ST	$(CAG)n_{II}$ expansion	[21]
$(CAG)_3 (CAA)_3 (CAG)_6 CAA CAG CAA (CAG)_{28} CAA CAG$	45	1 gen. (2 sibling)	ST	$(CAG)n_{II}$ expansion	[22]
$(CAG)_3 (CAA)_3 (CAG)_6 CAA CAG CAA (CAG)_{31} CAA CAG$	48	2 gen. (3 family members)	ST	$(CAG)n_{II}$ expansion	
$(CAG)_3 (CAA)_3 (CAG)_6 CAA CAG CAA (CAG)_{30} CAA CAG$	47	1 gen. (3 siblings)	ST	$(CAG)n_{II}$ expansion	
$(CAG)_3 (CAA)_3 (CAG)_9 CAA CAG CAA (CAG)_{16} CAA CAG CAA (CAG)_{16} CAA CAG$	55	1 gen. (2 siblings)	ST	partial duplication	
$(CAG)_3 (CAA)_3 (CAG)_n CAA CAG$	53–55	2 gen. (mother and 2 siblings)	NST	partial deletion + exp.	[23]
$(CAG)_3 (CAA)_3 (CAG)_9 CAA CAG CAA (CAG)_{31} CAA CAG$	51	single case	-	$(CAG)n_{II}$ expansion	
$(CAG)_3 (CAA)_3 (CAG)_9 CAA CAG CAA (CAG)_{23} CAA CAG$	43	single case	-	$(CAG)n_{II}$ expansion	
$(CAG)_3 (CAA)_3 (CAG)_9 CAA CAG CAA (CAG)_{30} CAA CAG$	50	single case	-	$(CAG)n_{II}$ expansion	[24]
$(CAG)_3 (CAA)_3 (CAG)_9 CAA CAG CAA (CAG)_{28} CAA CAG$	48	2 gen. (mother and 6 siblings)	ST	$(CAG)n_{II}$ expansion	
$(CAG)_3 (CAA)_3 (CAG)_9 CAA CAG CAA (CAG)_{24} CAA CAG$	44	single case	-	$(CAG)n_{II}$ expansion	
$(CAG)_3 (CAA)_3 (CAG)_{11} CAA CAG CAA (CAG)_{24} CAA CAG$	46	2 gen. (father and daughter) *	ST	$(CAG)n_{II}$ expansion	[25]
$(CAG)_3 (CAA)_3 (CAG)_9 CAA CAG CAA (CAG)_{27} CAA CAG$	47 [#]	single case	-	$(CAG)n_{II}$ expansion	[26]
$(CAG)_3 (CAA)_3 (CAG)_6 CAA CAG CAA (CAG)_{31} CAA CAG$	48 [#]	single case	-	$(CAG)n_{II}$ expansion	
$(CAG)_3 (CAA)_3 (CAG)_9 CAA CAG CAA (CAG)_{44} CAA CAG$	53	1 gen. (2 siblings)	ST	$(CAG)n_{II}$ expansion	
$(CAG)_3 (CAA)_3 (CAG)_9 CAA CAG CAA (CAG)_{57} CAA CAG$	53–66	2 gen. (father and son)	NST	$(CAG)_{nII}$ expansion	[27]
$(CAG)_3 (CAA)_3 (CAG)_8 CAA CAG CAA (CAG)_{26} CAA CAG$	45	1 gen. (2 siblings) **	ST	$(CAG)n_{II}$ expansion	[28]
$(CAG)_3 (CAA)_3 (CAG)_n CAA CAG$	53–58	2 gen. (father and son)	NST	partial deletion + exp	[29]

Table 1. *Cont.*

CAG/CAA Repeat Structure in TBP Gene	Repeats No.	Analyzed Cases	Stability	Possible Mutation Mechanism	Reference
$(CAG)_3$ $(CAA)_3$ $(CAG)_9$ CAA CAG CAA $(CAG)_{25}$ CAA CAG	45	1 gen. (2 siblings) **	-	$(CAG)n_{II}$ expansion	[30]
$(CAG)_3$ $(CAA)_3$ $(CAG)_8$ CAA CAG CAA $(CAG)_{26}$ CAA CAG	45	1 gen. (2 siblings) **	-	$(CAG)n_{II}$ expansion	
$(CAG)_3$$(CAA)_3$ $(CAG)_9$ CAA CAG CAA $(CAG)_{15}$ CAA CAG CAA $(CAG)_{17}$ CAA CAG	55	single case	-	partial duplication + exp	[31]
$(CAG)_3$ $(CAA)_3$ $(CAG)_9$ CAA CAG CAA $(CAG)_{32}$ CAA CAG	52	1 gen. (2 cousins)	ST	partial duplication + exp	[32]
$(CAG)_3$ $(CAA)_3$ $(CAG)_n$ CAA CAG	49–53	2 gen. (father and 2 siblings)	NST	partial deletion + exp	[33]
$(CAG)_3$ $(CAA)_4$ $(CAG)_n$ CAA CAG	53–66	2 gen. (father and daughter)	NST	partial deletion + exp	[34]
$(CAG)_3$ $(CAA)_4$ $(CAG)_n$ CAA CAG	54–55, 61	2 gen. (father and 2 daughters)	NST	partial deletion + exp	
$(CAG)_3$ $(CAA)_4$ $(CAG)_n$ CAA CAG	51–51, 52	2 gen. (father and 2 daughters)	NST	$(CAG)n_{II}$ expansion	
$(CAG)_3$ $(CAA)_3$ $(CAG)_9$ CAA CAG CAA $(CAG)_{28}$ CAA CAG	45	single case	-	$(CAG)n_{II}$ expansion	[35]
$(CAG)_3$ $(CAA)_3$ $(CAG)_n$ CAA CAG	50–55	2 family members	NTS	partial deletion + exp	[11]
$(CAG)_3$$(CAA)_3$ $(CAG)_8$ CAA CAG CAA $(CAG)_{35}$ CAA CAG	54	single case (de novo, paternal transmission)	-	$(CAG)_{nII}$ expansion	[36]
$(CAG)_3$ $(CAA)_3$ $(CAG)_{12}$ CAA CAA $(CAG)_{13}$ CAA $(CAG)_{16}$ CAA CAG	52	2 gen. (father and proband)	?	probably rearrangement pat. expanded repeats	[37]
$(CAG)_3$ $(CAA)_3$ $(CAG)_9$ CAA CAG CAA $(CAG)_{23}$ CAA CAG	43	single case	-	$(CAG)_{nII}$ expansion	
$(CAG)_3$ $(CAA)_3$ $(CAG)_{30}$ CAA CAG CAA $(CAG)_{16}$ CAA CAG	57	single case	-	$(CAG)_{nI+nII}$ expansion	
$(CAG)_3$ $(CAA)_3$ $(CAG)_9$ CAA CAG CAA $(CAG)_{35}$ CAA CAG	55	single case	-	$(CAG)_{nII}$ expansion	
$(CAG)_3$ $(CAA)_3$ $(CAG)_9$ CAA CAG CAA $(CAG)_{31}$ CAA CAG	51	single case	-	$(CAG)_{nII}$ expansion	
$(CAG)_3$ $(CAA)_3$ $(CAG)_9$ CAA CAG CAA $(CAG)_{25}$ CAA CAG	45	2 gen. (mother and proband)	-	$(CAG)_{nII}$ expansion	
$(CAG)_3$ $(CAA)_3$ $(CAG)_9$ CAA CAG CAA $(CAG)_{16}$ CAA CAG CAA $(CAG)_{16}$ CAA CAG	55	1 gen. (2 siblings)	ST	partial duplication	[38]
$(CAG)_3$ $(CAA)_3$ $(CAG)_{11}$ CAA CAG CAA $(CAG)_{27}$ CAA CAG	47	2 gen. (mother and son)	ST	$(CAG)n_{II}$ expansion	
$(CAG)_3$ $(CAA)_3$ $(CAG)_9$ CAA CAA AGG $(CAG)_3$ $(CAA)_3$$(CAG)_9$ CAA CAG CAA $(CAG)_{18}$ CAA CAG	56	single case (de novo paternal transmission)	ST	partial duplication	

ST—stable during transmission. NST—unstable during transmission,—stability unknown (single cases), ?—stability unknown no data of probands' parents, gen.—generation; *—only daughter's allele was sequenced, we can assume the stable transmission as in both cases the number of the repeats was exactly the same; **—only one sibling's allele was sequenced, we can assume the stable transmission as in both cases the number of the repeats was exactly the same; #—patient was homozygous for the expanded allele according to repeat number and structure of the region.
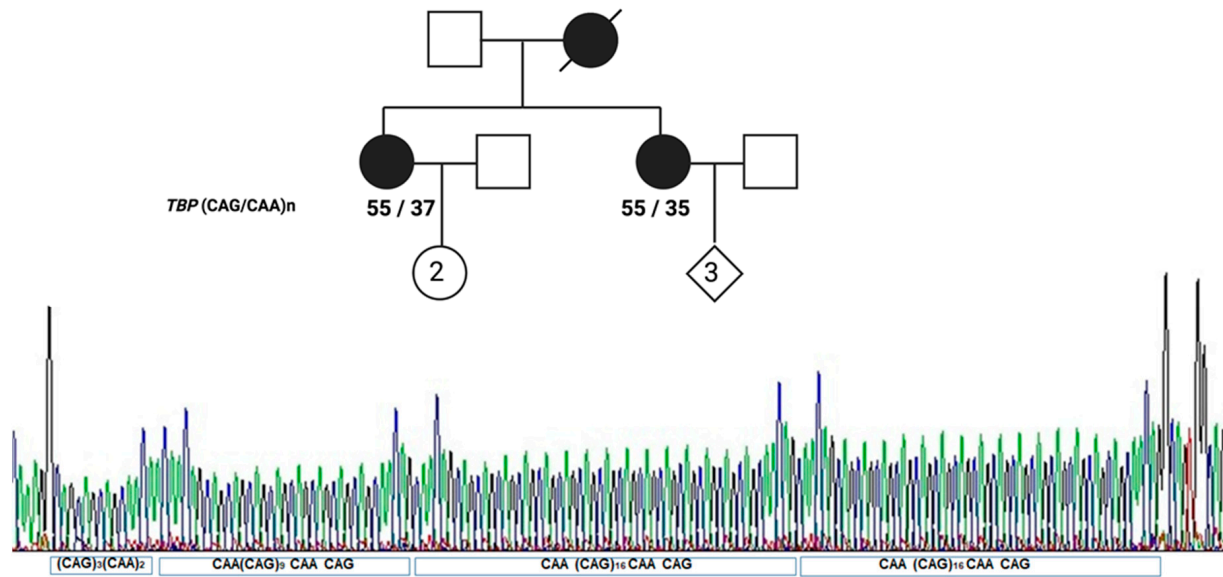
Figure 1. Pedigree and sequence of the CAG>GAA repeat region in the *TBP* gene of a family with two patients with the same number of repeats in the pathogenic allele $n = 55$ and the same repeat configuration. The sequence is given below the chromatogram, where the domain division is also marked according to the scheme we propose. (Created with BioRender.com; accessed on 15 July 2024).

Table 2. Comparison of the accepted and newly proposed CAA/CAG repeat motifs' organization in the *TBP* gene. The possible influence of the motifs' organization on the mutation mechanism and regional stability during transmission (based on already published cases listed in Table 1).

CAG/CAA Motif Organization in *TBP* Gene	
$(CAG)_3 (CAA)_3$ $(CAG)_{6-11}$ CAA CAG CAA $(CAG)_{9-21}$ CAA CAG I II III IV V	Five-unit organization of the CAG/CAA repeats region according to Gostout et al., 1993 [18]. Two polymorphic $(CAG)_n$ tracks are as follows: **II**—less variable $n = 6$–11 ($n = 9$; 91,7%), **IV**—$n = 9$–21 ($n = 14$–17; 83%).
$(CAG)_3 (CAA)_2$ $CAA(CAG)_{6-11}CAA$ CAG $CAA(CAG)_{9-21}CAA$ CAG I II III	Proposed 3-unit organization of CAG/CAA repeat region. Two motifs of identical structure II and III with variable $(CAG)_n$ repeat number. Similar to the five-unit model for a possible mechanism of mutation—expansion of the $(CAG)_n$ in unit III as a main mechanism or rearrangements within the region as a result of the interaction (secondary structure formation) between the whole motifs.
Possible model of mutation in identified CAG/CAA motifs	
$[(CAG)_3 (CAA)_2] [CAA (CAG)_9 CAA CAG] [CAA (CAG)_{>28} CAA CAG]$	Expansion of the $(CAG)_n$ track in the second $CAA(CAG)_nCAACAG$ motif due to its hairpin formation and polymerase slippage—slow, multistep process alleles rather stable in transmission no loss of the basic configuration.
$[(CAG)_3 (CAA)_2] [CAA (CAG)_{45} CAA CAG]$	Expansion of the $(CAG)_n$ track as a consequence of its instability after deletion of one of the $CAA(CAG)_nCAACAG$ motifs, probably as a result of the hairpin structure formation on the leading strand. Less stable in transmission (in analyzed data max. increased CAG repeats number +7)
$[(CAG)_3 (CAA)_2] [CAA (CAG)_9 CAA CAG] [CAA (CAG)_{16} CAA CAG]_2$ $[(CAG)_3 (CAA)_2] [CAA (CAG)_9 (CAA)_3 (CAG)_9 CAA CAG]_2 [CAA (CAG)_{19} CAA CAG]$ $[(CAG)_3 (CAA)_2] CAA (CAG)_{12} (CAA)_2 (CAG)_{13} [CAA (CAG)_{16} CAA CAG]$	Partial duplication of the CAG/CAA repeats region due to the formation of the FEN1 resistant structures on the 5′ flap end of the Okazaki fragment or by triplex structure forming between the Okazaki fragment and template strand.

3. Discussion

SCA17 is a typical polyQ neurodegenerative disorder, but on the other hand, unique due to the organization of the polyQ-coding region in the *TBP* gene. The pure polyglutamine track is coded by CAG stretches interrupted by CAA repeats (both coding Gln/Q). As was mentioned, Gostout et al. [18] proposed the organization of these repeats as a five-unit complex [(CAG)$_3$ (CAA)$_3$] [(CAG)nI] [CAA CAG CAA] [(CAG)nII] [CAA CAG] with two possible configurations according to Gao et al., 2008, as type I and type II, with and without domain III [CAA CAG CAA] [11]. Stretches of type I are more stable somatically and germinally, and expansion usually occurs here in domain IV [CAG]n but more complex rearrangements are also observed, including the partial duplications of domain III—[(CAG)$_3$ (CAA)$_3$] [(CAG)$_9$] [CAA CAG CAA (CAG)$_{16}$ CAA CAG CAA (CAG)$_{16}$] [CAA CAG]—but also more complicated rearrangements including partial duplication [(CAG)3 (CAA)$_3$] [(CAG)$_9$] [CAA CAG CAA (CAG)$_9$] [(CAA)$_3$ (CAG)$_9$ CAA CAG CAA (CAG)$_{19}$] [CAA CAG] or deletions [(CAG)3 (CAA)3] [(CAG)n] (Table 1). Type II may be a derivative of the original five-component repeat combination, which was created by deletions of domain III [(CAG)$_3$ (CAA)$_3$] [(CAG)n] CAA CAG], which destabilizes the sequence of domain II repeats and can easily expand in inheritance parent-offspring.

Here, based on the results from previous studies as well as our own data, a different scheme of the mutable CAG/CAA repeats' organization in the *TBP* gene is proposed, not as a five-unit [(CAG)$_3$ (CAA)$_3$] [(CAG)nI] [CAA CAG CAA] [(CAG)nII] [CAA CAG], but as a three-unit tract [(CAG)$_3$(CAA)2]$_I$ [CAA (CAG)n CAA CAG]$_{II}$ [CAA (CAG)n CAA CAG]$_{III}$ with different mutable components, two of which are identical except for the (CAG)n repeats number (Table 2). The (CAG)n tract in segment II is rather stable, with nine repeats in most cases; the main unstable (CAG)n tract is in the last segment (III). The suggested motifs' organization indicates three possible ways of repeat region elongation: (a) Expansion of the second (CAG)n domain (Gostout's unit IV), with its stable transmission in families, as the presence of two repeats of the [CAA (CAG)n CAA CAG] motif stabilizes the whole region and only allows for a slow multistep expansion of the second (CAG)n track. (b) Partial deletion of the region, probably units II and III, followed by the strong expansion in the remaining (CAG)n domain, which is also unstable during transmission, and (c) the partial duplication of the repeats region as a result of the duplication of the unit III or more complicated rearrangements due to the formation of FEN-1 resistant structures during replication.

To date, in the case of a partial duplication, it was not so easy to determine which parts of the region are prone to mutation; the model of repeats organization proposed here may better explain changes involved in the partial deletion or duplication of this region. In the [CAA (CAG)n CAA CAG] segments, even in the normal range, the (CAG)n repeats a number in most cases (>80%) and is long enough to form a hairpin structure (the number of repeats required for expansion/contraction has been estimated as 15/17, respectively) [35]. The loss of one of such motifs destabilizes the repeats' structure and the remaining (CAG)n track is more prone to expansion by easier formation of the secondary structures during replication and causing the polymerase slippage. Although only two such cases have been published, they clearly show the instability in the transmission leading to the anticipation of the disease inheritance [30,34]. Koide et al. proposed a hypothetical model of de novo expansion in the *TPB* gene by 5′-flap single-strand generation in the Okazaki fragment and formation of the FEN-1 resistant structure (hairpin) which was finally duplicated during replication [17]. The formation of such structures could also be the base of partial deletions when the hairpin at the 5′-flap end of the Okazaki fragment is recognized and processed by FEN-1 nuclease. However, this is probably not the case since FEN-1 is able to process the 5′ flap DNA of CAC/CTG repeats in a length-dependent manner. It has been demonstrated in vitro that human FEN-1 nuclease cuts such molecules containing up to 21 repeats but its activity decreases when the flap has over 11 repeats, and this may be correlated with a growing tendency to form higher structures suppressing FEN-1 activity [40]. A possibly removed hairpin, formed between the whole motif II and

part of motif III, would be at the threshold value for such a reaction (about 21 repeats). This might rather suggest a single case of contraction, as a result of the hairpin formation on the lagging strand template, followed in the next generations by (CAG)n expansion in a single and unstable [CAA (CAG)n CAA CAG] motif. Two cases with partial duplication of the repeats' region, published by Nakamura et al. [22], and the second one presented here also show the stable transmission of such mutated alleles. In both families, the presence of the duplication of the second [CAA (CAG)n CAA CAG] motif (unit III) was reported in siblings. In both pairs, the structure of repeats was identical among them. This may suggest that their affected parents—the father in the case of Nakamura et al. and the mother in the authors' case—were carriers of identical duplication, and the presence of an additional [CAA (CAG)$_{16}$ CAA CAG] motif did not destabilize the whole region during transmission (Table 2). As in the case of Koide et al. [17] this mutation could also be described based on the model of the repeat expansion via an FEN-1-resistant flap formation [7]. There are two possible ways in which such structures form within the CAA(CAG)$_{16}$CAA CAG unit: a hairpin, backflap, or triplex structure between this motif and the preceding repeat unit.

The possible models of CAG/CAA repeat expansion in the *SCA17/TBP* gene discussed above are based on the generally accepted replication models of expansion as the best way of explaining human trinucleotide instability. But as it has also been suggested, the formation of the ssDNA secondary structures and their resolution may involve different processes such as the excision of the secondary structure, repair at double-strand breaks by the recombinational mechanisms (gene conversion repair) or end-joining, mismatch repair, or gap repair [5]. The involvement of a particular mechanism depends on the cell cycle or differentiation phase as well as on the age and the cell type. Interestingly, FEN-1 seems to be involved in all those cellular processes as one of the crucial enzymes maintaining genome stability [8].

4. Materials and Methods

The search terms for the literature search in PubMed up to 7 July 2024, were as follows: [ataxia]; [TATA-binding protein]; [CAG expansion] OR [TBP]; [ataxia]; and [CAG/CAA expansion].

In total, 81 records were found; 2 were removed before screening because they were not in English. Among the 79 records assessed for eligibility, 2 studies were a review, 53 publications were not relevant as they did not contain the full sequence of the TBP abnormal allele and no patient data related to the sequence, and 10 considered only animal or cellular models.

In total, 64 records were excluded from further screening.

In total, 17 studies included in the manuscript contained a detailed human sequence of abnormal TBP alleles, 4 reports were additionally included in the analysis based on references in Zuhlke and Burk, 2007 [33], and 1 was our results [38].

Finally, 19 publications (78 cases) were included in this study; the references positions were [17,18,21–38].

In the case of 3 patients diagnosed at the Institute of Psychiatry and Neurology, procedures were performed according to EMQN protocols, including *TBP* PCR, as described by Nakamura et al. 2001 [22]. To establish the sequence of the CAG/CAG repeats region, PCR products were cloned in the pJET1.2 vector using the A/T Cloning system [Fermentas], and inserts were sequenced by the Sanger method using the BigDye Terminatorv.3.1 sequencing standard kit [Applied Biosystem] and pJET1.2 forward and reverse sequencing primers.

5. Conclusions

SCA17 is inherited in an autosomal dominant manner, and the offspring of affected individuals are at a 50% risk of inheriting the expanded *TBP* allele. The molecular diagnosis of this disorder is established by the identification of an abnormal CAG/CAA repeat expansion in the TBP gene and determining the number of repeats. However, the age of onset, severity, specific symptoms, and progression of the disease is variable and cannot

be precisely predicted by family history or the size of expansion. If it can be generally assumed that a higher number of repetitions correlates with an earlier age of onset (late vs. juvenile form of SCA17), its instability, unlike diseases caused by pure repeat tracts, depends on CAG/CAA configurations. The presence of CAA interruptions breaks up the repetitive sequence into shorter homogenous triplet tracts. This may have a stabilizing influence during DNA replication and could reduce strand slippage. There is no apparent segregation of particular phenotypic traits with CAG/CAA repeat tract structures or repeat lengths; however, it was shown that more complex motif rearrangements are identified in less typical cases.

This is why not only the repeats number but also a detailed analysis of the CAG/CAA repeat structure in expanded alleles should not be performed as it may have a prognostic value for affected families—stability/instability during transmission and possible anticipation of the disease.

Author Contributions: Conceptualization D.H.-Z. and A.S.; genetic analysis D.H.-Z. and A.S.; writing, original draft preparation, review and editing A.S. and D.H.-Z. All authors have read and agreed to the published version of the manuscript.

References

1. Depienne, C.; Mandel, J.L. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *Am. J. Hum. Genet.* **2021**, *108*, 764–785. [CrossRef] [PubMed]
2. Mitas, M. Trinucleotide repeats associated with human disease. *Nucleic Acids Res.* **1997**, *25*, 2245–2253. [CrossRef] [PubMed]
3. Cummings, C.J.; Zoghbi, H.Y. Fourteen and counting: Unraveling trinucleotide repeat diseases. *Hum. Mol. Genet.* **2000**, *9*, 909–916. [CrossRef]
4. Sinden, R.R.; Potaman, V.N.; Oussatcheva, E.A.; Pearson, C.E.; Lyubchenko, Y.L.; Shlyakhtenko, L.S. Triplet repeat DNA structures and human genetic disease: Dynamic mutations from dynamic DNA. *J. Biosci.* **2002**, *27*, 53–65. [CrossRef] [PubMed]
5. Lenzmeier, B.A.; Freudnreich, C.H. Trinucleotide repeat instability: A hairpin curve at the crossroads of replication, recombination and repair. *Cytogenet. Genome Res.* **2003**, *100*, 2–24. [CrossRef] [PubMed]
6. E Pearson, C.; Sinden, R.R. Trinucleotide repeat DNA structures: Dynamic mutations from dynamic DNA. *Curr. Opin. Struct. Biol.* **1998**, *8*, 321–330. [CrossRef] [PubMed]
7. Gordenin, D.A.; Kunkel, T.A.; Resnik, A.M. Repeat expansion—All in a flap? *Nat. Genet.* **1997**, *16*, 116–118. [CrossRef] [PubMed]
8. Liu, Y.; Kao, H.I.; Bambra, R.A. Flap endonuclease 1: The central component of DNA metabolism. *Annu. Rev. Biochem.* **2004**, *73*, 589–615. [CrossRef] [PubMed]
9. Liu, Y.; Zang, H.; Veeraraghavan, J.; Bambra, R.; Fruedenreich, C.H. Saccharomyces cerevisiae Flap endonuclease 1 uses flap equi-libration to maintain triplet repeat stability. *Mol. Cell Biol.* **2004**, *24*, 4049–4064. [CrossRef]
10. E Pearson, C. Slipping while sleeping? Trinucleotide repeat expansions in germ cells. *Trends Mol. Med.* **2003**, *9*, 490–495. [CrossRef]
11. Gao, R.; Matsuura, T.; Coolbaugh, M.; Zühlke, C.; Nakamura, K.; Rasmussen, A.; Siciliano, M.J.; Ashizawa, T.; Lin, X. Instability of ex-panded CAG/CAA repeats in spinocerebellar ataxia type 17. *Eur. J. Hum. Genet.* **2008**, *16*, 215–222. [CrossRef] [PubMed]
12. White, R.J.; Jackson, S.P. The TATA-binding protein: A central role in transcription by RNA polymerases I, II and III. *Trends Genet.* **1992**, *8*, 248–288. [CrossRef]
13. Chung, M.Y.; Ranum, L.P.; Duvick, L.A.; Servadio, A.; Zoghbi, H.Y.; Orr, H.T. Evidence for a mechanism predisposing to intergenera-tional CAG repeat instability in spinocerebellar ataxia type I. *Nat. Genet.* **1993**, *5*, 254–258. [CrossRef] [PubMed]
14. Costanzi-Porrini, S.; Tessarolo, D.; Abbruzzese, C.; Liguori, M.; Ashizawa, T.; Giacanelli, M. An interrupted 34-CAG repeat SCA-2 allele in patients with sporadic spinocerebellar ataxia. *Neurology* **2000**, *54*, 491. [CrossRef] [PubMed]
15. Chong, S.S.; Almqvist, E.; Telenius, H.; LaTray, L.; Nichol, K.; Bourdelat-Parks, B.; Goldberg, Y.P.; Haddad, B.R.; Richards, F.; Sillence, D.; et al. Contribution of DNA Sequence and CAG Size to Mutation Frequencies of Intermediate Alleles for Huntington Disease: Evidence from Single Sperm Analyses. *Hum. Mol. Genet.* **1997**, *6*, 301–309. [CrossRef] [PubMed]

16. Bauer, P.; Laccone, F.; Rolf, A. Trinucleotide repeat expansion in SCA17/TBP on white patients with Huntington's disease-like phenotype. *J. Med. Genet.* **2004**, *41*, 230–232. [CrossRef] [PubMed]

17. Koide, R.; Kobayashi, S.; Shimohata, T.; Ikeuchi, T.; Maruyama, M.; Saito, M.; Yamada, M.; Takahashi, H.; Tsuji, S. A neurological disease caused by an expanded CAG trinucleotide repeat in the TATA-binding protein gene: A new polyglutamine disease? *Hum. Mol. Genet.* **1999**, *8*, 2047–2053. [CrossRef] [PubMed]

18. Gostout, B.; Liu, Q.; Sommer, S.S. "Criptic" repeating triplets of purines and pirimidines (cRRY(i)) are frequent and polymorphic: Analysis of coding cRRY(i) in the proopiomelanocortin (POMOC) and TATA-binding protein (TBP) genes. *Am. J. Hum. Genet.* **1993**, *52*, 1182–1190. [PubMed]

19. Magri, S.; Nanetti, L.; Gellera, C.; Sarto, E.; Rizzo, E.; Mongelli, A.; Ricci, B.; Fancellu, R.; Sambati, L.; Cortelli, P.; et al. Digenic inheritance of STUB1 variants and TBP polyglutamine expansions explains the in-complete penetrance of SCA17 and SCA48. Digenic inheritance of STUB1 variants and TBP polyglutamine expansions explains the incomplete penetrance of SCA17 and SCA48. *Genet. Med.* **2022**, *24*, 29–40. [CrossRef]

20. Barbier, M.; Davoine, C.S.; Petit, E.; Porché, M.; Guillot-Noel, L.; Sayah, S.; Fauret, A.L.; Neau, J.P.; Guyant-Maréchal, L.; Deffond, D.; et al. Intermediate repeat expansions of TBP and STUB1: Genetic modifier or pure digenic inher-itance in spinocerebellar ataxias? *Genet. Med.* **2023**, *25*, 100327. [CrossRef]

21. Fujigasaki, H.; Martin, J.-J.; De Deyn, P.P.; Camuzat, A.; Deffond, D.; Stevanin, G.; Dermaut, B.; Van Broeckhoven, C.; Dürr, A.; Brice, A. CAG repeat expansion in the TATA box-binding protein gene causes autosomal dominant cerebellar ataxia. *Brain* **2001**, *124*, 1939–1947. [CrossRef]

22. Nakamura, K.; Jeong, S.-Y.; Uchihara, T.; Anno, M.; Nagashima, K.; Nagashima, T.; Ikeda, S.-I.; Tsuji, S.; Kanazawa, I. SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Hum. Mol. Genet.* **2001**, *10*, 1441–1448. [CrossRef]

23. Zühlke, C.; Hellenbroich, Y.; Dalski, A.; Hagenah, J.; Vieregge, P.; Riess, O.; Klein, C.; Schwinger, E. Different types of repeat expansion in the TATA-binding protein gene are associated with a new form of inherited ataxia. *Eur. J. Hum. Genet.* **2001**, *9*, 160–164. [CrossRef] [PubMed]

24. Silveira, I.; Miranda, C.; Guimarães, L.; Moreira, M.C.; Alonso, I.; Mendonça, P.; Ferro, A.; Pinto-Basto, J.; Coelho, J.; Ferreirinha, F.; et al. Trinucleotide repeats in 202 families with ataxia: A small expanded (CAG)n allele at the SCA17 locus. *Arch. Neurol.* **2002**, *59*, 623–629. [CrossRef]

25. Zühlke, C.; Gehlken, U.; Hellenbroich, Y.; Schwinger, E.; Bürk, K. Phenotypical variability of expanded alleles in the TATA-binding protein gene. *J. Neurol.* **2003**, *250*, 161–163. [CrossRef] [PubMed]

26. Stevanin, G.; Fujigasaki, H.; Lebre, A.S.; Camuzat, A.; Jeannequin, C.; Dode, C.; Takahashi, J.; San, C.; Bellance, R.; Brice, A.; et al. Huntington's disease-like phenotype due to trinucleotide repeat expansions in the TBP and JPH3 genes. *Brain* **2003**, *126*, 1599–1603. [CrossRef]

27. Zühlke, C.H.; Spranger, M.; Spranger, S.; Voigt, R.; Lanz, M.; Gehlken, U.; Hinrichs, F.; Schwinger, E. SCA17 caused by homozygous repeat expansion in TBP due to partial isodisomy 6. *Eur. J. Hum. Genet.* **2003**, *11*, 629–632. [CrossRef]

28. Maltecca, F.; Filla, A.; Castaldo, I.; Coppola, G.; Fragassi, N.; Carella, M.; Bruni, A.; Cocozza, S.; Casari, G.; Servadio, A.; et al. Intergenerational instability and marked anticipation in SCA-17. *Neurology* **2003**, *61*, 1441–1443. [CrossRef] [PubMed]

29. Toyoshima, Y.; Yamada, M.; Onodera, O.; Shimohata, M.; Inenaga, C.; Fujita, N.; Morita, M.; Tsuji, S.; Takahashi, H. SCA17 homozygote showing Huntington's disease-like phenotype. *Ann. Neurol.* **2004**, *55*, 281–286. [CrossRef]

30. Brusco, A.; Gellera, C.; Cagnoli, C.; Saluto, A.; Castucci, A.; Michielotto, C.; Fetoni, V.; Mariotti, C.; Migone, N.; Di Donato, S.; et al. Molecular genetics of hereditary spinocerebellar ataxia: Mutation analysis of spinocerebellar ataxia genes and CAG/CTG repeat expansion detection in 225 Italian families. *Arch. Neurol.* **2004**, *61*, 727–733. [CrossRef]

31. Shatunov, A.; Fridman, E.; Pagan, F.; Leib, J.; Singleton, A.; Hallett, M.; Goldfarb, L. Small *de novo* duplication in the repeat region of the TATA-box-binding protein gene manifest with a phenotype similar to variant Creutzfeldt-Jakob disease. *Clin. Genet.* **2004**, *66*, 496–501. [CrossRef] [PubMed]

32. Bruni, A.C.; Takahashi-Fujigasaki, J.; Maltecca, F.; Foncin, J.F.; Servadio, A.; Casari, G.; D'adamo, P.; Maletta, R.; Curcio, S.A.M.; De Michele, G.; et al. Behavioral Disorder, Dementia, Ataxia, and Rigidity in a Large Family With TATA Box-Binding Protein Mutation. *Arch. Neurol.* **2004**, *61*, 1314–1320. [CrossRef] [PubMed]

33. Zühlke, C.; Bürk, K. Spinocerebellar ataxia type 17 is caused by mutations in the TATA-box binding protein. *Cerebellum* **2007**, *6*, 300–307. [CrossRef] [PubMed]

34. Rasmussen, A.; De Biase, I.; Fragoso-Benítez, M.; Macías-Flores, M.A.; Yescas, P.; Ochoa, A.; Ashizawa, T.; Alonso, M.E.; Bidichandani, S.I. Anticipation and intergenerational repeat instability in spinocerebellar ataxia type 17. *Ann. Neurol.* **2007**, *61*, 607–610. [CrossRef] [PubMed]

35. Watanabe, M.; Monai, N.; Jackson, M.; Yamamoto-Watanabe, Y.; Ikeda, Y.; Suzuki, C.; Tomiyama, M.; Kawarabayashi, T.; Kimura, T.; Seino, Y.; et al. A Small Trinucleotide Expansion in the TBP Gene Gives Rise to a Sporadic Case of SCA17 with Abnormal Putaminal Findings on MRI. *Intern. Med.* **2008**, *47*, 2179–2182. [CrossRef] [PubMed]

36. Bech, S.; Petersen, T.; Nørremølle, A.; Gjedde, A.; Ehlers, L.; Eiberg, H.; Hjermind, L.E.; Hasholt, L.; Lundorf, E.; Nielsen, J.E. Huntington's disease-like and ataxia syndromes: Identification of a family with a de novo SCA17/TBP mutation. *Parkinsonism Relat. Disord.* **2010**, *16*, 12–15. [CrossRef] [PubMed]

37. Ivanova, E.; Nuzhnyi, E.; Abramycheva, N.; Klyushnikov, S.; Fedotova, E.; Illarioshkin, S. Mutation analysis of the TATA box-binding protein (TBP) gene in Russian patients with spinocerebellar ataxia and Huntington disease-like phenotype. *Clin. Neurol. Neurosurg.* **2022**, *222*, 107473. [CrossRef] [PubMed]
38. Sulek, A.; Hoffman-Zacharska, D.; Zdzienicka, E.; Empel, J.; Zaremba, J. SCA17—A rare form of spinocerebellar ataxia identified in Poland. *Eur. J. Hum. Genet.* **2002**, *10* (Suppl. S1), 269.
39. Lee, S.; Park, M.S. Human FEN-1 can process the 5′-flap DNA of CTG/CAG triplet repeat derived from human genetic diseases by length and sequence dependent manner. *Exp. Mol. Med.* **2002**, *34*, 313–317. [CrossRef]
40. Hemricksen, L.A.; Tom, S.; Liu, Y.; Bambara, R.A. Inhibition of Flao Endonuclease 1 by Flap Secondary structure and relevance to repeat sequence expansion. *J. Biol. Chem.* **2000**, *275*, 16420–16427. [CrossRef]