Analysis of Strand Slippage in DNA Polymerase Expansions of CAG/CTG Triplet Repeats Associated with Neurodegenerative Disease*

(Received for publication, September 29, 1997, and in revised form, December 17, 1997)

John Petruska, Michael J. Hartenstine, and Myron F. Goodman

From the Department of Biological Sciences, Hedco Molecular Biology Laboratories, University of Southern California, Los Angeles, California 90089-1340

L lengthy expansions of trinucleotide repeats are found in DNA of patients suffering severe neurodegenerative age-related diseases. Using a synthetic self-priming DNA, containing CAG and CTG repeats implicated in Huntington’s disease and several other neurological disorders, we measure the equilibrium distribution of hairpin folding and generate triplet repeat expansions by polymerase-catalyzed extensions of the hairpin folds. Expansions occur by slippage in steps of two CAG triplets when the self-priming sequence (CTG)$_{16}$(CAG)$_{4}$ is extended with proofreading-defective Klenow fragment (KF exo$^{-}$) from Escherichia coli DNA polymerase I. Slippage by two triplets is 20 times more frequent than by one triplet, in accordance with our finding that hairpin loops with even numbers of triplets are 1–2 kcal/mol more stable than their odd-numbered counterparts. By measuring triplet repeat expansion rates as they evolve over time, individual rate constants for expansion from 4 to 18 CAG repeats are obtained. An empirical expression is derived from the data, enabling the prediction of slippage rates from the ratio of hairpin CTG/CTG interactions to CAG/CTG interactions. Slippage is initiated internally in the hairpin folds in preference to melting inward from the 3’ terminus. The same triplet expansions are obtained using proofreading-proficient KF exo$^{+}$, provided 10–100-fold higher dNTP concentrations are present to counteract the effect of 3’-exonucleolytic proofreading.

The human genome has an abundance of simple sequence repetitions that are unstable and tend to expand in large numbers in some genetic loci. A prime example is the CAG/CTG class of triplet repeats whose large expansions occur in genes associated with Huntington’s disease and six other neurological disorders (1). Such expansions represent a novel form of inheritance, as mentioned in the previous study of self-priming with triplet repeats (6). Strand slippage can occur when there is local strand separation in DNA regions containing tandem repetitions (9). For example, in a region of CAG/CTG repeats, local separation creates single-strand loops of CAGs and CTGs whose repetitive character may allow the two strands to be displaced (i.e. slipped) by an integral number of triplets. Such slippage in the presence of DNA polymerases can result in the addition of integral numbers of triplets to give triplet repeat expansions (2–8). Large loops that form stable hairpin structures are expected to increase the probability of slippage and expansion (8, 10–14).

We (14) and others (10, 12, 13, 15) have shown that single strands of CTG repeats form stable hairpin structures as a result of base pairing between antiparallel CTGs, yielding G-C and C-G base pairs alternating with T-T mismatches. Similar hairpins also form in strands of CAG repeats but are not as stable (14–16), probably because A-A mismatches are bulkier and more destabilizing than T-T (17). Our thermal denaturation studies of strand folding (14) indicated that hairpin stability increases more slowly than expected with increasing numbers of repeats. A plausible explanation is that strands with more repeating triplets have more degrees of freedom and tend to form several shorter hairpin loops instead of one long one. For example, a strand with 10 triplet repeats may fold into a single hairpin structure, whereas a longer strand with 30 repeats may form a more complex structure with two or three hairpin folds of similar size (14).

Evidence of strand slippage in CAG/CTG repeat regions during DNA polymerization has been obtained using various gel-based primer extension assays (5–8). Our assay differs from earlier ones mainly in the priming event. We engineered a DNA oligonucleotide of triplet repeats so that it is primed by intramolecular folding, rather than by intermolecular association. Being independent of concentration, the self-priming remains efficient at low DNA concentrations, unlike intermolecular priming that requires higher DNA concentrations. By using low concentrations, we avoid uncertainties arising from multimeric associations between DNA strands. Also, our starting primer 3’-end is fully complementary to template repeats, instead of partially complementary as in the only previous study of self-priming with triplet repeats (6).

In the present study we use a self-priming strand containing a large number of CTGs relative to CAGs, to measure the influence of CTG/CTG interactions on strand slippage in polymerase-catalyzed expansions of correctly paired (CAG/CTG) repetitions. This approach provides for the first time an opportunity to measure slippage and expansion rates as a function of the number of interacting triplets, for DNA polymerases with and without proofreading 3’-exonucleolytic activity.

EXPERIMENTAL PROCEDURES

Materials

DNA Polymerases—Escherichia coli DNA polymerase I Klenow fragment (KF) mutant, exo$^{-}$ (D355A,E357A), devoid of proofreading 3’-exonucleolytic activity, was purified from overproducing strains (18). Normal proofreading-proficient KF polymerase (exo$^{+}$) was obtained

1 The abbreviations used are: KF, Klenow fragment; A, self-priming DNA sequence, 5’-(CTG)$_{16}$(CAG)$_{4}$-3’; A$_{9}$, initial hairpin fold of sequence A, having n CTG triplets at the 5’-end available as template for extending the primer 3’-end by n CAG triplets; A$_{9}$ + (CAG)$_{n}$, blunt-end hairpin formed by extending A$_{9}$ by n CAG triplets with DNA polymerase.

© 1998 by The American Society for Biochemistry and Molecular Biology, Inc.
commercially (U. S. Biochemical Corp./Amersham Corp.).

Oligonucleotide Synthesis—The self-primer DNA 60-mer, (CTG)(n=)
(CAG)₄, and 30-mer (CTG)₅(CAG)₄ were synthesized by an Applied
Biosystems DNA/RNA synthesizer, using β-cyanoethyl phosphoramidites,
and purified by denaturing polyacrylamide gel electrophoresis. The
90-mer DNA marker (CTG)₉(CAG)₄, was purchased from Operon
Technologies Inc. and obtained lyophilized after high pressure liquid
chromatography purification. Purified DNA samples were dialyzed ex-
tensively against a low ionic strength buffer (5 mM NaH₂PO₄, 5 mM
NaHPO₄, 1 mM Na₃EDTA, pH 7.0) and stored at -70 °C.

Methods

Melting Analysis—Thermal denaturation profiles were obtained for the
60- and 30-mer DNAs at the same strand concentration (1.9 μM) in
low ionic strength buffer, by measuring UV absorbance A₂₆₀ versus
temperature, while raising temperature from 25 to 85 °C at a constant
rate of 2 °C/min.

End Labeling and Equilibration—Samples of the self-primer 60-
mer used in polymerase reactions were 5’-end-labeled with [3²P] using
γ-[³²P]ATP and T4 polynucleotide kinase (U. S. Biochemical Corp./
Amersham Corp.) in kinase buffer (50 mM Tris-HCl, pH 7.6, 10 mM
MgCl₂, and 2 mM mercaptoethanol). Labeled samples at a strand
concentration of 100 nM were heat-denatured at 100 °C for 5 min and
allowed to refold by cooling at room temperature. The results ob-
tained in polymerase reactions showed no dependence on the rate of
cooling, as expected for intramolecular folding. The labeled strands
were stored at 4 °C to avoid any potential intermolecular associations
promoted by freezing (14).

Extension Reactions—Radiolabeled DNA at 10 nM strand concen-
tration was incubated 5 min at 37 °C in reaction buffer (50 mM Tris-HCl,
ph 7.5, 10 mM MgCl₂, 1 mM dithiothreitol, and 0.05 mM mg/ml bovine serum albumin) to allow equilibration of DNA structure. Typically, 60 μl of
dNA solution was then micropipetted into 15 μl of enzyme + dNTP
template solution (buffer) held in a microcentrifuge tube at 37 °C, at which
point running time (t) for reaction was started. Final enzyme concen-
tration was approximately 60 nM in each case; the dNTPs used (N=C,
A, G) were in equimolar amounts ranging from 0.1 to 10 μM. At times
indicated (t = 0.5 min, etc.) a 5-μl aliquot of reaction mixture was
removed and quenched with 25 μl of 20 mM EDTA + 20 μM formamide.

Denaturing Gel Electrophoresis—Reaction products were separated
into bands according to product size, by denaturing gel electrophoresis
at 2000 V on a 12% polyacrylamide (40 cm x 40 cm x 0.2 mm) with 16
formamide as denaturant, in TBE buffer (89 mM Tris borate, pH 8.3, 2
mM Na₂EDTA). Gels dried on paper were scanned with Molecular
Dynamics Storm 860 PhosphorImager, and band intensities were inte-
grated by FragmentNT analysis software. Each band intensity was evalu-
ated as percent of total integrated band intensities in the correspond-
ing lane.

Expansion Rate Analysis—The 60-mer sequence (Aₐ) forms hairpin
loops (Aₐ) with even and odd numbers of bases in the hairpin (Fig. 1).
In Fig. 1, the loops of DNA polymerase and dNTPs, the lengths are rapidly
extended (in seconds) to blunt-ended forms, Aₐ + (CAG)₄, seen as gel
band intensities (Fig. 1), we find that the

RESULTS

The self-primer sequence, Aₐ = (CTG)₆(CAG)₄, forms a se-
ries of hairpin loops (Fig. 1), to which integral numbers of CAG
triplets can be added by polymerase-catalyzed reaction with appro-
priate dNTP substrates (N = C, A, G). In Fig. 1, A₀
represents the most stable loop in the form of a blunt-ended
hairpin; A₁ corresponds to slippage of A₀ by a single triplet, A₂
slippage by two triplets; etc. Initially, there is an equilibrium
distribution of loops reflecting their thermodynamic stability in
the absence of polymerase. We verify that stable folds are
formed by examining thermal denaturation profiles obtained
by plotting UV absorbance (A₂₆₀) versus temperature (Fig. 2).

When DNA polymerase and dNTPs are added, the original
loops are extended to blunt ends in a matter of seconds, so that
the rates at which blunt-end hairpins rearrange by slippage
and expand by triplet additions can be accurately measured
over a period of minutes. To establish the equilibrium amounts
of individual hairpins and their expansion rates, we analyzed
polymerase-catalyzed reaction products separated into discrete
disks by electrophoresis on a denaturing, formamide-polyacryl-
amide gel. The products are first analyzed as a function of reaction
time (t) at physiologically low (0.1 to 1 μM) dNTP concentration,
using a proofreading-deficient DNA polymerase, Klenow fragment exo-. The results shown (Fig. 3) are obtained

FIG. 1. Possible hairpin folds of self-primer 60-mer DNA se-
quence (CTG)₆(CAG)₄. The even-numbered loops (A₀, A₁, etc.) and
less stable odd-numbered loops (A₂, A₃, etc.) are extendable by even
and odd numbers of CAG triplets, respectively. The solid dots indicate
correct base pairs; the asterisk indicates the 5’-end radiolabeled with
[³²P]P. The most stable loop is blunt-ended A₀, having the most
correct base pairs, namely 23. The loops indicated as much less stable have
less 4 CAGs correctly paired with complementary antiparallel CTGs.
Upon reaction with DNA polymerase and dNTPs (N = C, A, G), the
initial loops A₀ are rapidly extended (in seconds) to blunt-end products,
A₁ + (CAG)₄, seen as band intensities (n = 0, 1, 2, etc.) by denaturing
gel electrophoresis. With increasing reaction time, the changes in I₄
indicate the rates of strand slippage in the blunt-ended hairpins
and further expansion by addition of more CAG triplets.

by a triplet (I₄ → I₅), converting the even-numbered loop A₀ to the
odd-numbered loop A₁ (Fig. 1). The assumption is that I₄ → I₅ occurs
with rate constant k₄, whereas I₅ → I₆ occurs with rate constant k₅
for 2-triplet slippage. When applying the corresponding differential
equation, dI₄/dt = k₄I₄ - k₅I₅, we find that k₅' (for 1-triplet slippage)
is an order of magnitude less than k₄ (for 2-triplet slippage), whereas k₅
for 2-triplet slippage of an odd-numbered loop is consistent with k₄
and k₅ for 2-triplet slippage of even-numbered loops.

Even-Numbered Loop | Odd-Numbered Loop
|-----------------|-----------------|
| A₀: C₅G(TAG)₆(CAG)₄ | A₁: G(CAG)₅(CAG)₄ |
| C₇G(TAG)₆(CAG)₄* | C₇G(TAG)₆(CAG)₄* |
| A₂: C₈G(TAG)₆(CAG)₄ | A₃: G(CAG)₈(CAG)₄ |
| G₈TAG(TAG)₈(CAG)₄* | G₈TAG(TAG)₈(CAG)₄* |
| A₄: C₉G(TAG)₆(CAG)₄ | A₅: G(CAG)₉(CAG)₄ |
| G₉TAG(TAG)₉(CAG)₄* | G₉TAG(TAG)₉(CAG)₄* |
| A₆: C₁₀G(TAG)₆(CAG)₄ | A₇: G(CAG)₁₀(CAG)₄ |
| G₁₀TAG(TAG)₁₀(CAG)₄* | G₁₀TAG(TAG)₁₀(CAG)₄* |
| A₈: C₁₁G(TAG)₆(CAG)₄ | A₉: G(CAG)₁₁(CAG)₄ |
| G₁₁TAG(TAG)₁₁(CAG)₄* | G₁₁TAG(TAG)₁₁(CAG)₄* |
| A₁₀: C₁₂G(TAG)₆(CAG)₄ | A₁₁: G(CAG)₁₂(CAG)₄ |
| G₁₂TAG(TAG)₁₂(CAG)₄* | G₁₂TAG(TAG)₁₂(CAG)₄* |

Much Less Stable:

A₁₂: C₁₃G(TAG)₆(CAG)₄ |
| A₁₃: G(CAG)₁₃(CAG)₄ |
A₁₄: C₁₄G(TAG)₆(CAG)₄ |
| A₁₅: G(CAG)₁₄(CAG)₄ |

Downloaded from http://www.jbc.org/ by guest on July 23, 2018
with 0.2 μM each of the three dNTP substrates (N = C, A, G). This concentration appears high enough to obtain near-maximum velocity of extension with a minimum of artifacts such as terminal transferase activity, seen at higher [dNTP] as an increase in background “pause” bands (data not shown).

**Thermal Denaturation Profiles**—Upon heat denaturation (Fig. 2), the 60-mer (CTG)$_{16}$(CAG)$_{4}$ shows two sigmoidal transitions, with melting temperature $T_m = 56$ and 80°C at low ionic strength. By comparison, the 50-mer (CTG)$_{15}$(CAG)$_{4}$, having 10 fewer CTG repeats, shows only the higher transition with $T_m$ near 80°C. The biphasic melting curve of the 60-mer is consistent with the kind of hairpin folding shown in Fig. 1, in which the 4 CAG triplets are correctly base-paired with 4 CTG repeats to give the more stable component melting at 80°C, whereas the remaining 12 CTG triplets are more weakly paired to give the less stable component melting at 56°C. The latter $T_m$ value is close to that observed for hairpin folding of CTG triplets alone, e.g., (CTG)$_{10}$ and (CTG)$_{20}$ each with $T_m = 51°C$ (14). However, the melting curve is unable to resolve the kinds of hairpin folds present. The various possible folds (Fig. 1) initially have the same strand length, but when extended with polymerase they acquire increasing lengths, which can be separated by electrophoresis on formamide-polyacrylamide gel.

**Initial Hairpin Loops**—As seen in Fig. 1, starting at the 5'-end marked by *, the 60-mer sequence A has a potential template of 16 CTG repeats followed by a primer of 4 CAG repeats. Upon hairpin folding, so that the CAGs align with CTGs in antiparallel fashion, the (CAG)$_{4}$ primer can be correctly base-paired with any 4 successive antiparallel CTGs, shown as (GTC)$_{4}$ in Fig. 1. The first loop, A$_{10}$, is a blunt-ended structure with 3'- and 5'-ends juxtaposed. This is the most stable fold because it allows remaining CTGs to form the maximum number of base pairs with each other.

The even-numbered loops A$_{2}$, A$_{4}$, ..., A$_{10}$ belong in the same group as A$_{10}$, because they each have the same kind of 4-base hairpin bend and an even number of overhanging triplets. The odd-numbered loops A$_{1}$, A$_{3}$, ..., A$_{11}$ belong in a separate group, because they each have one more base pair stabilizing the hairpin bend. For completeness, much less stable loops, A$_{12}$ to A$_{50}$, having fewer than 4 correctly paired CAG/CTG combinations, are also shown (Fig. 1).

Being blunt-ended, A$_{10}$ is not suitable for extension by DNA polymerases until it rearranges by strand slippage into one of the other possible forms with an overhanging 5'-template, (GTC)$_{n}$* (Fig. 1). These forms, A$_{n}$ (n = 1, 2, etc.), are easily extended to blunt-end products, A$_{n} + (CAG)$_{4}, by the addition of (CAG)$_{4}$ opposite (GTC)$_{n}$* . The products are resolved as a series of bands on a denaturing gel, with band intensity $I_n$ proportional to the amount of product, (A$_{n} + (CAG)$_{4})*. The 5'-end (*) is radiolabeled with $^{32}$P to obtain quantitative band intensity measurements with a PhosphorImager.

**Loop Resolution by Electrophoretic Analysis of Polymerase-catalyzed Extension Products**—Using proofreading-deficient DNA polymerase KF exo -, we observed complete extension of initial loops within 0.5 min of reaction (Fig. 3). A 6-fold excess of polymerase is used to saturate all the self-primed loops originally present so that they can be completely extended to blunt ends in seconds, before appreciable primer/template slippage can occur. Also, we use a dNTP concentration low enough (0.2 μM) to avoid the formation of anomalous pause bands arising by base misinsertion or by blunt end addition of a single nucleotide (19).

In Fig. 3 it is clear that the most intense bands correspond to...
additions of even numbers of CAG triplets. The bands formed by adding 0, 2, 4, etc. triplets are much more intense than those formed by adding 1, 3, etc. triplets. This striking result shows that even-numbered bands (Fig. 1) are significantly more abundant, i.e. more stable, than odd-numbered, in agreement with earlier observations (10). The way the band intensities change with reaction time, indicates the rates at which the blunt-end loops (Fig. 1) are significantly more abundant over slippage by one triplet. As Fig. 1 shows, 

$$I_0 \rightarrow I_2 \rightarrow I_4 \rightarrow I_6 \rightarrow I_8 \rightarrow I_{10} \rightarrow I_{12} \rightarrow I_{14}$$

FIG. 5. Kinetic scheme used to evaluate rate constants for slippage and expansion by two triplets. The rate constants $k_n$ (n = 0, 2, 4, etc.) indicate the rate at which band intensity $I_n$ is converted to $I_{n+2}$, as a result of 2-triplet slippage and polymerase-catalyzed expansion by 2 CAG triplets.

FIG. 4. Plot of band intensities versus reaction time for the first three main bands representing extensions by 0, 2, and 4 CAG triplets. Band intensity $I_n$ indicates the amount of the original blunt-ended loop $A_0$ (Fig. 1), and $I_2$ and $I_4$ indicate the amounts of the extension products, $A_2 + (CAG)_2$ and $A_4 + (CAG)_2$, respectively. Each band intensity is expressed as percent of total integrated band intensities in the corresponding lane (Fig. 3), measured by PhosphorImager.

By extrapolating to $t = 0$ we obtain $I_0(t \rightarrow 0)$, the initial amount of hairpin $A_0$ (Table I). Similarly, using $dI_2/dt = k_2I_0 - k_2I_2$, we are able to evaluate $k_2$, relative to $k_0$, and by extrapolation to estimate the initial amounts of the even-numbered hairpins with $n = 4, 6, 8, 10, 12, 14$. In the case of $A_{12}$, the initial amount (0.5%) is barely above the minimum level of detection, and the rate constant for slippage is not determined (ND, Table I).

Because odd-numbered loops give much less intense bands, comparable to faint pause bands in the background, we can only evaluate initial amounts for $A_1$, $A_3$, and $A_5$ (Table I). In the case of $A_1$, there is sufficient intensity to estimate a rate constant $k_1$ for 2-triplet slippage ($I_1 \rightarrow I_3$), assuming $A_0$ contributes to $A_1$ ($I_0 \rightarrow I_1$) with rate constant $k_0$ for 1-triplet slippage. In this case, $dI_2/dt = k_0I_0 - k_1I_1$, when applied to our data, yields $k_0 = 0.04 \pm 0.02$ min$^{-1}$ and $k_1 = 0.7 \pm 0.1$ min$^{-1}$ (Table I, in brackets).

Relationship between Slippage Rate and Number of Interacting Triples—In Table I (last column) we see that the slippage rate constant $k_n$ decreases with increasing number, $n$, of added CAGs. The rate constant is highest (0.8 min$^{-1}$) for $n = 0$, the initial blunt-end hairpin $A_0$ (Fig. 1), which has the highest number of CTG/CTG interactions relative to CAG/CTG interactions. As Fig. 1 shows, $A_0$ has 4 strong CAG/CTG base pairing interactions and 5 weaker CTG/CTG interactions, so that (CTG/CTG)/(CAG/CTG) = 5/4. The longer blunt-end hairpins, $A_n + (CAG)_n$, formed by adding $n$ CAG triplets complementary to the template (CTG)$_n$, (Fig. 1), gain $n$ more CAG/CTG interactions while losing $n/2$ CTG/CTG interactions. Thus, while (CAG/CTG) = 4 + $n$ increases with $n$, both (CTG/CTG)/(CAG/CTG) = (10 - $n$)/(8 + 2$n$), decrease with $n$.

The $k_0$ value (on a logarithmic scale) shows a positive correlation with (CAG/CTG) = 5 - $n/2$ and a negative correlation with (CAG/CTG) = 4 + $n$ (Fig. 6a). In each case, the correlation is nonlinear for even values of $n$ from 0 to 8. However, when log $k_0$ is plotted against the ratio, (CTG/CTG)/(CAG/CTG) = (10 - $n$)/(8 + 2$n$), the points fall close to a straight line (Fig. 6b). Furthermore, $k_1$ obtained for 2-triplet slippage between odd-numbered loops (Table I, in brackets) also falls on this line, at the point corresponding to $n = 1$ (Fig. 6b, open square).

The equation of the line fitted by least squares (Fig. 6b) is

$$\log k_0 = -0.38x - 0.56,$$

where $x = (\text{CTG/CTG})/(\text{CAG/CTG})$. This equation can also be written as $k_0 = a \exp(-0.38x)$, where $a = 0.28$ min$^{-1}$ and $\beta = 0.87$. We can interpret $a$ as the slippage rate constant for CAG/CTG interactions alone and $\exp(-0.38x)$ as the factor arising from the influence of (CTG/CTG) on the slippage of (CAG/CTG). This factor appears to hold as long as slippage by 2 triplets does not change the character of the hairpin bend made by CAGs.

The even-numbered loops $A_{n}$ to $A_{14}$ (Fig. 1) have the same kind of 4-base bend made with 2 CTGs. However, in $A_{14}$ the bend acquires a CAG in place of a CTG at the expense of a CAG/CTG base-pairing interaction, making it more difficult for $A_{14}$ to $A_{12}$ slippage to occur. As seen in Table I (last column), the slippage rate constant drops dramatically, from 0.28 min$^{-1}$ for $k_{14}$ ($I_{14} \rightarrow I_{12}$) to 0.08 min$^{-1}$ for $k_{10}$ ($I_{10} \rightarrow I_{8}$). This marked reduction in slippage rate is also evident in the accumulation of band intensity $I_{10}$, corresponding to 10 CAG triplets added.
Initial amounts of hairpin folds in sequence $A = (\text{CTG})_m(\text{CAG})_n$ and corresponding slippage rate constants for 2-triplet expansions

The rate constants for slippage and expansion by 2 CAG triplets are evaluated by analyzing gel band intensities versus reaction time with DNA polymerase KF exo$^-$ and 0.2 $\mu$m dNTP ($N = C, A, G$) at 37 °C. Evaluations are made for each hairpin structure $A_i$, starting with $A_0$ (Fig. 1), assuming rapid extension (in seconds) to a blunt-state, $k_0$, and then slower expansion (in minutes) with rate constant $k_i$, because of slippage by 2 triplets as illustrated in Fig. 5. The initial amount of $A_0$ is found by analyzing band intensity $I_i$ versus $t$ and extrapolating to $t = 0$. A good fit is obtained assuming slippage by 2 triplets, so that $dI/dt = k_i I_0 - k_1 I_1$.

The rate constants for slippage and expansion by 2 CAG triplets are evaluated by analyzing gel band intensities versus reaction time with DNA polymerase KF exo$^-$ and 0.2 $\mu$m dNTP, band intensities similar to those found initially with KF exo$^-$ are observed in the 1st min of reaction, but expansion by slippage is greatly reduced because of 3′-end degradation with increasing time (Fig. 7). As dNTP concentration is increased to counteract the degradation, the band patterns for expansion by slippage are recovered for longer times. At 1.0 $\mu$m dNTP, the same patterns of band intensities as in Fig. 3 are observed for the first 10 min, but some decreases in large expansions (notably $I_{12}$ and $I_{14}$) are evident at longer times. At 10 $\mu$m, no significant decreases are evident for at least 60 min. Thus, to obtain the same degree of expansion in the presence of proofreading as in its absence, apparently 10 to 100 times as much dNTP is required.

**DISCUSSION**

Single DNA strands containing tandemly repeated CTG or CAG triplets form hairpin structures which, although not as stable as normal DNA duplexes, are nonetheless remarkably stable (10-15, 20). To investigate the influence that such secondary structures have on strand slippage in polymerase-catalyzed repeat expansions, we have constructed a strand sequence, (CTG)$_{16}$(CAG)$_4$, forming a series of self-primed hairpins (Fig. 1) to which repeats of CAG triplets can be added by DNA polymerases. This sequence provides an opportunity to measure the influence of hairpin CTG/CAG interactions on primer/template slippage in repeating CAG/CTG regions. By using a DNA polymerase devoid of exonuclease activity (KF exo$^-$), we find that initial hairpin loops (formed at equilibrium) are fully extended to blunt-ended forms in a matter of seconds, so that the rates at which blunt-end loops rearrange by slippage and add more repeats can be measured over a period of minutes to about 1 h (Fig. 3). With this construct, we observe triplet repeat expansions in a series of discrete slippage steps for which rate constants can be evaluated.

Of the various hairpins formed initially (Fig. 1), those with an even number of bases in the bend appear most stable since they are found in the largest amounts (Table I). The series of even-numbered loops from $A_0$ to $A_{10}$ account for over 90% of all folded structures. As expected, the most stable loop is the...
initiate slippage is much more likely to occur in the CTG/CTG region. To initiate slippage by 2 triplets, three CAG/CTG interactions need to be disrupted at the primer 3'-end, but only one next to the CTG/CTG region where two weaker CTG/CTG interactions are more easily disrupted. Slippage by 2 triplets is favored over slippage by 1 triplet, in accordance with even-numbered loops of CTGs being more stable than odd-numbered (Table I). The rate constant for 2-triplet slippage increases with CTG/CTG interactions but decreases with CAG/CTG interactions (Fig. 6A), so that the rate constant increases exponentially with the ratio of CTG/CTG to CAG/CTG interactions, as seen by the linear plot of log(rate constant) versus this ratio (Fig. 6B). Furthermore, in the presence of 3'-exonucleolytic proofreading (Fig. 7), as long as there are CTG/CTG interactions present (n = 0–8) and slippage rates are 0.28 min⁻¹ or higher (Table I), less than 1.0 μM dNTP concentrations are needed to restore normal slippage patterns, whereas 10 μM concentrations are needed when CTG/CTG interactions are absent (n = 12 and 14) and slippage rates fall to 0.08 min⁻¹ or lower (Table I).

Since our system is single-stranded and relatively short, it is not certain how many of the points above may apply to long double-stranded DNA molecules in vivo. Nevertheless, with these points in mind, we can propose a simple model (Fig. 8) indicating how triplet repeat expansion may occur when there is local strand separation in a region of triplet repeats. Such separation could arise, for example, by negative supercoiling (21) of the kind needed to activate transcription or replication.

Once separated in a region of triplet repeats, the strands may

---

**Fig. 7.** Band patterns obtained by extension with proofreading-proficient DNA polymerase KF (exo⁺) at three different dNTP concentrations. Under incubation conditions like those used for proofreading-deficient KF exo⁻ (Fig. 3), at low (0.2 μM) dNTP concentration, the self-priming sequence 32P(CTG)16(CAG)4 is extended to blunt-ended states in less than 1 min as before, but expansion by slippage is markedly reduced because of strand degradation by proofreading 3'-exonuclease activity. To counteract the degradation, much higher dNTP concentrations are needed. A 5-fold increase (1.0 μM) restores the pattern of expansion by slippage seen in Fig. 3 for the first 10 min; another 10-fold increase (10 μM) is needed to maintain the pattern for the full 60 min.

---

**Fig. 8.** Proposed model of strand slippage and repeat expansion in vivo suggested by results of current experiments. The repeating arrows symbolize sequence repeats such as triplets. After strand separation and possible loop migration, a nick (cleavage by endonuclease) on one strand allows the loop on the opposite side to be stretched so that the nicked strand can be extended by DNA polymerase (possibly proofreading-deficient polymerase β in eukaryotes) and sealed by DNA ligase.
fold on themselves to form hairpin loops that are initially opposite each other but because of repetition are able to migrate apart by strand slippage and remain separated for longer periods. While apart, if a nick (single-strand cleavage) occurs opposite one of the loops, that loop can stretch and become linear so that a DNA polymerase, perhaps a proofreading-deficient polymerase involved in DNA repair, may fill in the gap to create an expansion on one strand. After the gap is filled and sealed by ligase, a nick opposite the other loop can lead to expansion on the other strand. According to this model, single-strand loops that become separated and nicked in the presence of polymerase may behave like the simple hairpin system we have studied.

Our simple system enables one to evaluate the contributions that individual single-strand interactions make to strand slippage in duplex repeat regions. By using the sequence (CTG)$_{16}$(CAG)$_{4}$, we have made an evaluation of CTG/CTG interactions contributing to slippage and expansion of CAG/CTG repeats. Similarly, by using the sequence (CAG)$_{16}$(CTG)$_{4}$, a corresponding evaluation can be made for CAG/CTG interactions. Since CAG/CAG interactions are known to be weaker, forming less stable hairpin folds (14–16, 20), it is of interest to see how much difference this makes in the slippage rate constants. Previously it has been generally considered (2, 3, 8, 11, 12, 20) that the greater the stability of single-strand folding, the greater the probability of strand slippage and expansion in repeat regions. However, it is also possible that a difference in folding stability on opposite strands may be important in determining strand slippage and expansion rates. This may be a reason why repeats of triplets such as GAA/TTC, having only purines on one side and pyrimidines on the other, also expand to give diseases such as Friedrich’s ataxia (22).

REFERENCES
1. Ashley, C. T., Jr., and Warren, S. T. (1995) Annu. Rev. Genet. 29, 703–728
2. Sinden, R. R., and Wells, R. D. (1992) Curr. Opin. Biotechnol. 3, 612–622
3. Wells, R. D. (1996) J. Biol. Chem. 271, 2675–2678
4. Richards, R. I., and Sutherland, G. R. (1994) Nat. Genet. 6, 114–116
5. Schlotterer, C., and Tautz, D. (1992) Nucleic Acids Res. 20, 211–215
6. Behn-Krappa, A., and Doerfler, W. (1994) Hum. Mutat. 3, 19–24
7. Ji, J., Clegg, N. J., Peterson, K. R., Jackson, A. L., Laird, C. D., and Loeb, L. A. (1996) Nucleic Acids Res. 24, 2835–2840
8. Ohshima, K., and Wells, R. D. (1997) J. Biol. Chem. 272, 16798–16806
9. Levinson, G., and Gutman, G. A. (1987) Mol. Biol. Evol. 4, 203–221
10. Darlow, J. M., and Leach, D. R. F. (1986) Genetics 114, 825–832
11. Kang, S., Ohshima, K., Shimizu, M., Amirhaeri, S., and Wells, R. D. (1995) J. Biol. Chem. 270, 27014–27021
12. Gary, A. M., Goedlner, G.,Jurani, N., Macura, S., and McMurray, C. T. (1995) Cell 81, 535–540
13. Mitas, M., Yi, A., Dill, J., Kamp, T. J., Chambers, K. J., and Haworth, I. S. (1995) Nucleic Acids Res. 23, 1050–1059
14. Petruska, J., Arnheim, N., and Goodman, M. F. (1996) Nucleic Acids Res. 24, 1992–1998
15. Smith, G. K., Jie, J., Fox, G. E., and Gao, X. (1995) Nucleic Acids Res. 23, 4303–4311
16. Yu, A., Dill, J., and Mitas, M. (1995) Nucleic Acids Res. 23, 4055–4057
17. Aboul-ela, F., Koh, D., Tinoe, I. J., and Martin, F. H. (1988) Nucleic Acids Res. 13, 4811–4825
18. Derbyshire, V., Freemont, P. S., Sanderson, M. R., Beele, L., Friedman, J. M., Joyce, C. M., and Stitz, T. A. (1988) Science 240, 199–201
19. Clark, J. M. (1988) Nucleic Acids Res. 16, 8677–8686
20. Mitas, M. (1997) Nucleic Acids Res. 25, 2245–2253
21. Geiblib, R., Bacolla, A., and Wells, R. D. (1997) J. Biol. Chem. 272, 16793–16797
22. Cossee, M., Schmitt, M., Cambazano, V., Reutenauer, L., Moutou, C., Mandel, J.-L., and Koenig, M. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 7452–7457
