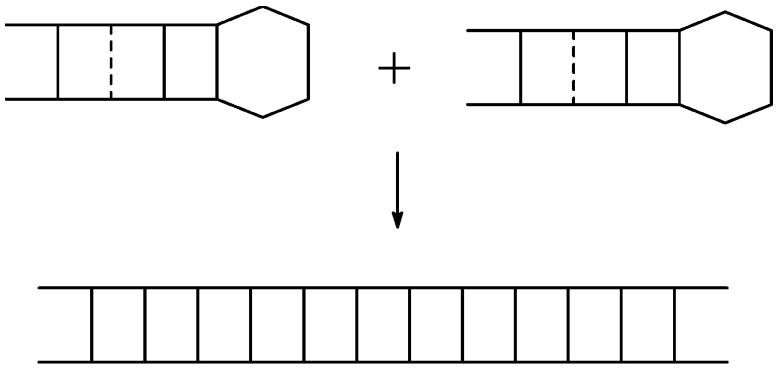
Article

The Influence of Sequence Context and Length on the Kinetics of DNA Duplex Formation from Complementary Hairpins Possessing (CNG) Repeats

Anthony M. Paiva, and Richard D. Sheardy

J. Am. Chem. Soc., **2005**, 127 (15), 5581-5585• DOI: 10.1021/ja043783n • Publication Date (Web): 23 March 2005



J|A|C|S
ARTICLES
Published on Web 03/23/2005

The Influence of Sequence Context and Length on the Kinetics of DNA Duplex Formation from Complementary Hairpins Possessing (CNG) Repeats

Anthony M. Paiva and Richard D. Sheardy*

Contribution from the Department of Chemistry and Biochemistry, Seton Hall University, 400 South Orange Avenue, South Orange, New Jersey 07079

Received October 12, 2004; E-mail: sheardri@shu.edu

Abstract: The formation of unusual structures during DNA replication has been invoked for gene expansion in genomes possessing triplet repeat sequences, CNG, where N = A, C, G, or T. In particular, it has been suggested that the daughter strand of the leading strand partially dissociates from the parent strand and forms a hairpin. The equilibrium between the fully duplexed parent:daugter species and the parent:hairpin species is dependent upon their relative stabilities and the rates of reannealing of the daughter strand back to the parent. These stabilities and rates are ultimately influenced by the sequence context of the DNA and its length. Previous work has demonstrated that longer strands are more stable than shorter strands and that the identity of N also influences the thermal stability [Paiva, A. M.; Sheardy, R. D. *Biochemistry* **2004**, *43*, 14218−14227]. Here, we show that the rate of duplex formation from complementary hairpins is also sequence context and length dependent. In particular, longer duplexes have higher activation energies than shorter duplexes of the same sequence context. Further, [(CCG):(GGC)] duplexes have lower activation energies than corresponding [(CAG):(GTC)] duplexes of the same length. Hence, hairpins formed from long CNG sequences are more thermodynamically stable and have slower kinetics for reannealing to their complement than shorter analogues. Gene expansion can now be explained in terms of thermodynamics and kinetics.

Certain genetic expansion diseases have been linked to the properties of trinucleotide repeat (i.e., triplet repeat) DNA sequences. Four of these sequences, CTG, CAG, and CCG/CGG, have been shown to be involved in the pathogenesis of myotonic dystrophy (DM), Huntington's disease (HD), and Fragile-X syndrome (FRAX), respectively.[1] Genes possessing $(CNG)_n$ sequences, where N = A, C, T, or G and n can range from 5 to 25 or more, can undergo novel mutational modes, termed "dynamic mutations," which lead to an expansion in the triplet repeat copy number, n, during DNA replication. The expansion in repeat copy number, however, is dependent upon the initial value of n. If n is below a certain threshold, which is sequence dependent, gene expansion will not occur. If n is above that threshold, gene expansion will occur and the disease state will be manifested.

There is no consensus mechanism of gene expansion in terms of the cellular events during replication. In the model proposed by the Wells group, the inherent flexibility of $(CNG)_n$ sequences can lead to an accumulation of superhelicity within those tracts.[2] Eventually the accumulated superhelical strain becomes so great that the DNA polymerase can no longer unwind double-stranded DNA downstream from the replication fork. Subsequent cellular events lead to dissociation of the polymerase complex followed by the formation of hairpin structures on the newly synthesized daughter DNA strand. If hairpin formation occurs prior to DNA polymerase reassociation, the polymerase will attach to a location upstream (3′ with respect to the parental strand position) of the dissociation point, leading to reiterative synthesis and expansion. The potential for gene expansion during replication ultimately depends on the fate of the hairpin formed on the daughter strand. The lifetime of the hairpin is dependent upon its thermodynamic stability and the kinetics of reassociation back to the parent strand. A long-lived hairpin will lead to reiterative synthesis, whereas a short-lived hairpin will eventually reassociate with the parental strand leading to no expansion in repeat number.

Triplet repeat DNA oligomers spontaneously form stable structures in vitro due to both the semi self-complementary sequence and the inherent strand flexibility, which allow alternative hydrogen-bonding arrangements of the resultant DNA secondary structure.[3] Thermodynamic analyses have led to model independent understanding of triplet repeat structural stabilities.[4] We have recently characterized the structure and thermodynamic stabilities of synthetic DNA oligomers with the general sequence of $(CNG)_n$ where N = A, C, T, or G and n = 4, 5, 10, 15, or 75.[5] All oligomers form stable unimolecular hairpin like structures at 25 °C in standard phosphate buffer (pH 7.0) with 115 mM Na^+. These hairpins possess stems with G:C base pairs and N:N mismatches and loops of 3−4 unpaired

(1) (a) Mandel, J.-L. *Nat. Genet.* **1994**, *4*, 453−455. (b) Pearson, C. E.; Sinden, R. R. *Biochemistry* **1996**, *35*, 5041−5053. (c) Reddy, P. S.; Housman, D. E. *Curr. Opin. Cell Biol.* **1997**, *9*, 364−372. (d) Gacy, A. M.; Geollner, G.; Juranic, N.; Macura, S.; McMurray, C. T. *Cell* **1995**, *81*, 553−540.
(2) (a) Gellibolian, R.; Bacolla, A.; Wells, R. D. *J. Biol. Chem.* **1997**, *272*, 16793−16797. (b) Bacolla, A.; Gellibolian, R.; Shimizu, M.; Amirhaeri, S.; Kang, S.; Ohshima, K.; Larson, J. E.; Harvey, S. C.; Stollar, B. D.; Wells, R. D. **1997**, *272*, 16783−16792.

10.1021/ja043783n CCC: $30.25

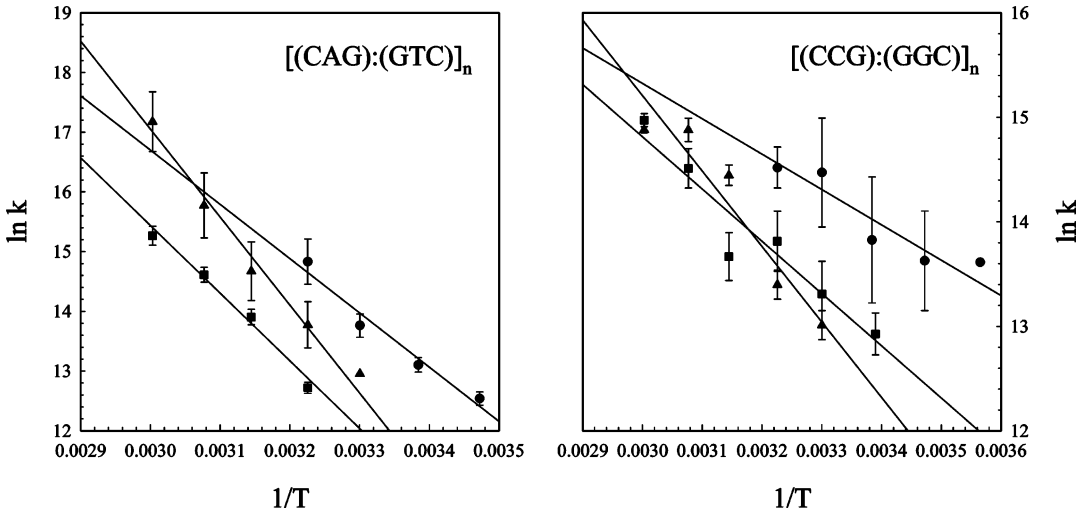


Figure 1. The Arrhenius plots (ln k vs $1/T$) for duplex formation of $[(CAG):(GTC)]_n$ and $[(CCG):(GGC)]_n$ for $n = 4$ (●), $n = 10$ (■), and $n = 15$ (▲). Points represent the average of two to four determinations, and the drawn lines represent the least squares linear fits. For these studies, the DNA oligomers (5×10^{-6} to 7×10^{-7} M in total strand) were prepared in standard phosphate buffer (10 mM total phosphate, 0.1 mM EDTA, pH 7.0 with NaCl added to a final $[Na^+]$ of 115 mM). The temperatures used for each duplex formed were chosen to be well below their melting temperatures.

Table 1. DNA Oligomers Used in This Study

general formula: $(CNG)_n$ N = A, T, C, or G and n = 4, 10, or 15			
$(CAG)_4$	$(CTG)_4$	$(CCG)_4$	$(CGG)_4$
$(CAG)_{10}$	$(CTG)_{10}$	$(CCG)_{10}$	$(CGG)_{10}$
$(CAG)_{15}$	$(CTG)_{15}$	$(CCG)_{15}$	$(CGG)_{15}$

Table 2. Thermodynamic Data for Duplex Formation

duplex	E_a (kcal/mol)	A (s^{-1})	ΔH^* (kcal/mol)	ΔS^* (cal/K mol)
$[(CAG):(GTC)]_4$	18.0 ± 0.5	1.2×10^{19}	17.5	27
$[(CAG):(GTC)]_{10}$	22.5 ± 0.5	2.8×10^{21}	22.0	38
$[(CAG):(GTC)]_{15}$	29.3 ± 0.5	3.8×10^{26}	28.8	61
$[(CCG):(GGC)]_4$	6.7 ± 0.5	1.2×10^{11}	6.2	−10
$[(CCG):(GGC)]_{10}$	9.9 ± 0.5	8.7×10^{12}	9.4	−1
$[(CCG):(GGC)]_{15}$	14.4 ± 0.5	1.0×10^{16}	13.9	13

bases. Further, the thermodynamic stabilities of these hairpins display both sequence context and length dependencies. In particular, hairpin stability order was found to increase as N = G > T > A > C. In addition, thermodynamic stabilities increase with increasing n, with gradual leveling off at $n > 45$ bases. Duplexes formed from complementary strands of equal length were found to be more stable than the component hairpins.

In the current study, we evaluate oligomer sequence and length dependence on the rate of duplex formation. The oligomers used are the same as those previously thermodynamically characterized (Table 1). All oligomers were synthesized and purified as previously described[5] and typically prepared in standard phosphate buffer (10 mM total phosphate, 0.1 mM EDTA, pH 7.0 with NaCl added to a final $[Na^+]$ of 115 mM). We have shown that all oligomers form hairpin secondary structures under the conditions used for these experiments.[5] To determine the rate constant for duplex formation from complementary hairpins, a particular oligomer was placed in the thermostated cuvette holder of a Cary 100 UV/vis spectrometer. The total strand concentration ranged from 5×10^{-6} to 7×10^{-7} M to obtain initial absorbances of 0.4−0.5. The complementary strand of the same length at the same total strand concentration, equilibrated to the same temperature, was added and mixed for 5 s. The decay in absorbance at 260 nm, starting at $t = 10$ s, was monitored as a function of time (i.e., the duplex has a lower extinction coefficient than either of the component strands). A typical decay can be seen in Figure 1S (Supporting Information). After $t = 15$ min, the absorbance no longer changes, suggesting equilibrium has been attained. The decay was best fit to a second-order reaction using the software of the Cary 100. The rates of duplex formation were slow enough to obtain good reproducibilities in rate constants (k) thus obtained.

Results and Discussion

The temperature dependence of the rate of duplex formation was evaluated by Arrhenius analysis according to eq 1:

$$\ln k = \ln A - E_a/RT \qquad (1)$$

where k is the observed rate constant at temperature T and R is the gas constant. As shown in Figure 1, linear dependencies were observed and the activation energies (E_a) and Arrhenius constants (A) were determined from the slopes and y intercepts, respectively. The line parameters for the Arrhenius plots are shown in Table 1S (Supporting Information).

The sequence context and length dependencies of the resolved kinetic parameters are indicated in Table 2. The activation energies for duplex formation for the (CAG):(CTG) series are higher than those of the (CCG):(CGG) series of the same length. The (CCG):(CGG) duplexes form more hydrogen bonds than the (CAG):(CTG) duplexes. Further, CC/GG and CG/GC base pair stacks are much more stable than CA/TG and AG/CT base pair stacks. Because (CCG):(CGG) duplexes form more hydro-

(3) (a) Zheng, M.; Huang, X.; Smith, G. K.; Yang, X.; Gao, X. *J. Mol. Biol.* **1996**, *264*, 323−326. (b) Mitas, M. *Nucleic Acids Res.* **1997**, *25*, 2245−2253. (c) Fry, M.; Loeb, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *33*, 4950−4954. (d) Nadel, Y.; Weisman-Shomer, P.; Fry, M. *J. Biol. Chem.* **1995**, *270*, 28970−28977. (e) Yu, A.; Barron, M. D.; Romero, R. M.; Christy, M.; Gold, B.; Dai, J.; Gray, D. M.; Haworth, I. S.; Mitas, M. *Biochemistry* **1997**, *36*, 3687−3699. (f) Chen, X.; Mariappan, S. V. S.; Catasti, P.; Ratliff, R.; Moyzis, R. K.; Laayoun, A.; Smith, S. S.; Bradbury, E. M.; Gupta, G. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5199−5203. (g) Mitas, M.; Yu, A.; Dill, J.; Kamp, T. J.; Chambers, E. J.; Haworth, I. S. *Nucleic Acids Res.* **1995**, *23*, 1050−1059. (h) Pearson, C. E.; Wang, Y.-H.; Griffith, J. D.; Sinden, R. R. *Nucleic Acids Res.* **1998**, *26*, 816−823.
(4) (a) Petruska, J.; Arnheim, N.; Goodman, M. F. *Nucleic Acids Res.* **1996**, *24*, 1992−1998. (b) Gacy, A. M.; McMurray, C. T. *Biochemistry* **1998**, *37*, 9426−9434. (c) Volker, J.; Makube, N.; Plum, G. E.; Klump, H. H.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14700−14705.
(5) Paiva, A. M.; Sheardy, R. D. *Biochemistry* **2004**, *43*, 14218−14227.

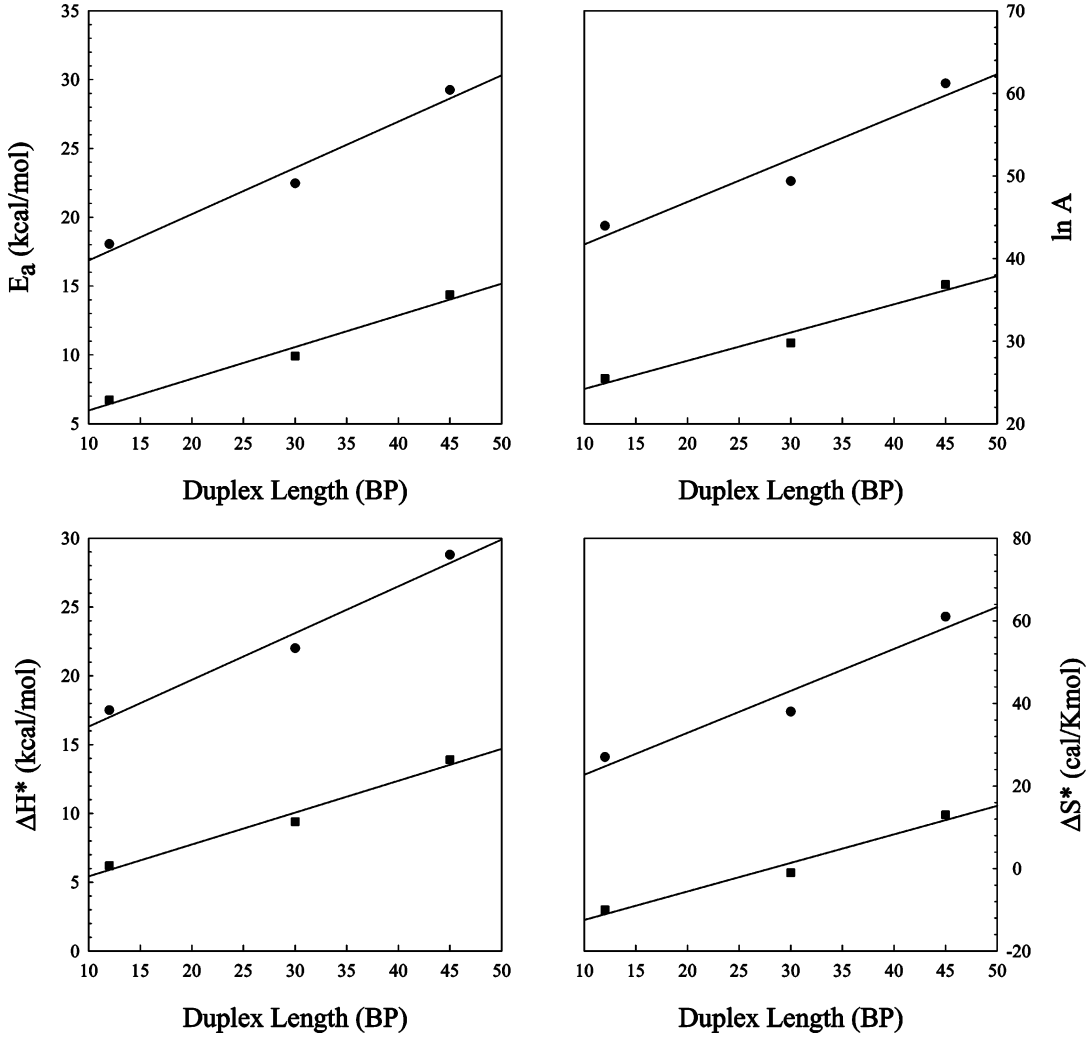


Figure 2. Correlation of kinetic parameters with duplex length for $[(CAG):(GTC)]_n$ (●) and $[(CCG):(GGC)]_n$ (■).

gen bonds and have more stable nearest neighbor interactions, they should require less nucleation free energies and therefore have lower activation energies. In addition, within a series, both E_a and ln A increase linearly with increasing length of duplex formed (Figure 2). Because the kinetic data are model dependent, it is likely that the dependence on length is governed by the cooperative unit length.

As noted above, the observed rate constants were obtained by mixing equimolar amounts of the complimentary hairpins; that is, no one strand was in excess. Because the duplexes formed are much more stable than the individual component hairpins from which they arise (see below), we assume that duplex formations are essentially complete after attaining equilibrium. Even in the unlikely event that duplex formation is not complete after equilibration, the relative trends reported in Table 1 would not be affected.

Using the transition state approach, the activation energy can be dissected into enthalpic and entropic components as follows:

$$\Delta H^* = E_a - RT \tag{2}$$

$$\Delta S^* = \mathrm{R}[\ln(Ah/k_bT) - 1] \tag{3}$$

where T is a reference temperature (310 K), R is the gas constant, h is Planck's constant, and k_b is Boltzman's constant. The values

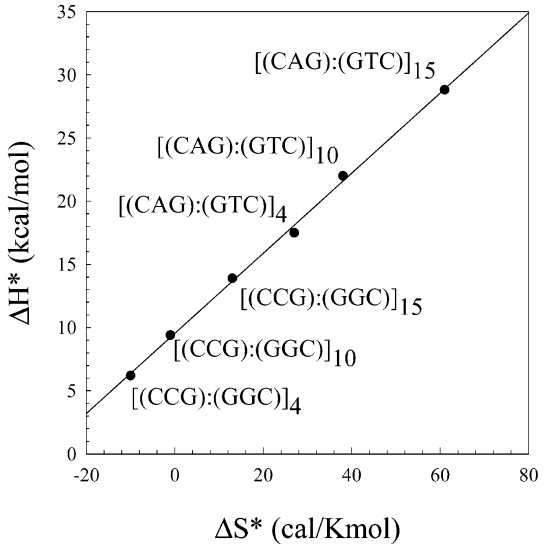


Figure 3. The activation enthalpy−entropy compensation for duplex formation of $[(CAG):(GTC)]_n$ and $[(CCG):(GGC)]_n$. Individual data points are identified, and the drawn line represents the least-squares fit.

for the thus calculated activation enthalpies and entropies are listed in Table 2. Clearly, duplex formation is enthalpically controlled with enthalpies increasing linearly with increasing

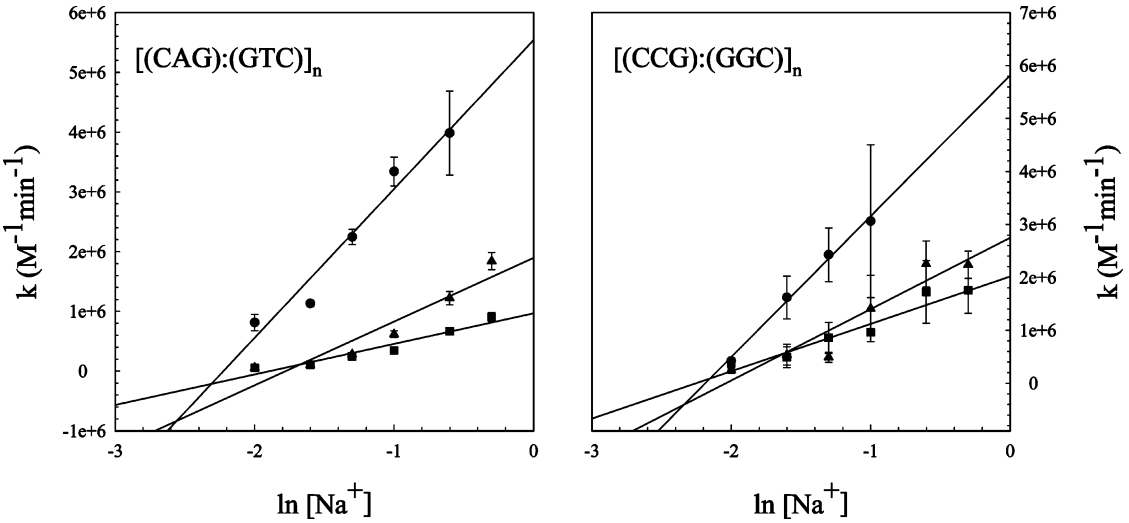


Figure 4. Plots of k versus ln[Na^+] for duplex formation at 20 °C of $[(CAG):(GTC)]_n$ and $[(CCG):(GGC)]_n$ for $n = 4$ (●), $n = 10$ (■), and $n = 15$ (▲). Points represent the average of two to four determinations, and the drawn lines represent the least squares linear fits. For these studies, the DNA oligomers (5×10^{-6} to 7×10^{-7} M in total strand) were prepared in standard phosphate buffer (10 mM total phosphate, 0.1 mM EDTA, pH 7.0) with NaCl added to a final [Na^+] ranging from 115 to 800 mM.

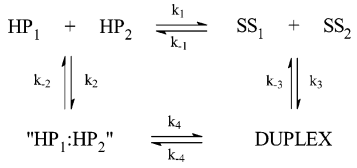
duplex length (Figure 2). Compensating the increasing enthalpies are entropies that also increase linearly with increasing duplex length (Figure 2). Further, enthalpies and entropies are higher for the [(CAG):(CTG)] series than the corresponding [(CCG):(CGG] series. This enthalpy:entropy compensation is clearly seen in Figure 3.

Our results are similar to those of Gacy and McMurray who investigated the kinetics of duplex formation for $[(CAG):(GTC)]_{25}$, $[(CGG):(GCC)]_{25}$, and $[(CAG):(GTC)]_{10}$ in that the longer duplexes were slower to anneal.[4b] Further, the magnitudes of our rate constants (10^5-10^6 M^{-1} min^{-1}, depending upon sequence, length, and ionic strength) are similar to their measured values. However, in contrast, we find that the [(CGG):(GCC)] duplexes form faster than the [(CAG):(GTC)] duplexes. In addition, the Breslauer group found that duplexes arising from X(CAG)X and X(GCT)X hairpins (where X = $(GC)_3$) formed only at elevated temperatures,[4c] whereas our duplexes formed quite readily at temperatures below the melting temperatures of the hairpins. In the case of the Breslauer hairpins, the 6 GC base pairs at the end of the hairpin stem stabilize the hairpin to a large degree such that the hairpin is kinetically trapped for duplex formation in the presence of its complement. Finally, the values for our activation energies are quite similar to those reported for duplex formation from complementary DNA hairpins of similar length by the Armitage group.[6] As with those studies, the secondary structures of the hairpins slow duplex formation. In our particular case, the kinetics were slow enough that they could be determined without the use of stop-flow techniques, that is, on the order of minutes.

The influence of the sodium ion concentration on the kinetics was also examined (Figure 4). Rate constants increase with increasing [Na^+] with the greatest dependence noted for the shortest duplexes. The duplex formed from any complementary pair will have a higher charge density than the component hairpins, thus mandating the uptake of Na^+ from the bulk solution upon duplex formation. The observed increase in rate constant with increasing [Na^+] is consistent with Na^+ uptake.

(6) Kushon, S. A.; Jordan, J. P.; Seifert, J. L.; Nielson, H.; Nielson, P. E.; Armitage, B. A. *J. Am. Chem. Soc.* **2001**, *123*, 10805−10813.

Scheme 1. Equilibria Involved with Duplex Formation from the Complementary Oligonucleotides[a]



[a] HP stands for HP, SS stands for for single strand, and HP:HP represents the "kissing" intermediate formed.

Clearly, the rates are more sensitive with respect to length with increasing [Na^+] than to sequence context. Longer strands experience more strand:strand repulsion due to their higher charge densities, so it is not surprising that longer stands have slower association kinetics under any ionic strength.

Scheme 1 shows the proposed mechanism for duplex formation from the individual hairpins. One pathway is the total denaturation of the hairpins (k_1) to single strands followed by annealing to the duplex (k_3). The other pathway involves formation of an intermediate "kissing" complex (k_2) followed by duplex formation (k_4). Becuase the kinetic experiments were carried out at temperatures well below the melting temperatures of the hairpins, the first pathway is less likely due to the low population of single-stranded oligonucleotides. Thus, the most probable pathway is the bimolecular association of two hairpins to form a "kissing" intermediate. Due to end fraying and the potential for slipped structures[5] for the longer oligomers, unpaired bases would be exposed at the ends of the hairpins. The "kissing" intermediate would thus form from the association of the unpaired bases of one hairpin with the unpaired bases of the complementary hairpin. The observed second-order kinetics of the reaction indicate that the formation of the "kissing" intermediate is the rate determining step, that is, $k_4 \gg k_2$.

The proposed scheme can also be used as a thermodynamic cycle to evaluate the free energy of duplex formation from complementary hairpins of equal length. We have previously reported the free energies for the hairpin to single-strand transition and the duplex to single-strand transition (Table 2S in Supporting Information).[5] As can be seen by the values reported in Table 3, the hairpin to duplex transition is highly

Table 3. Free Energy of Duplex Formation from Complementary Hairpins of Equal Length[a]

duplex	ΔG_1 (kcal/mol)	ΔG_3 (kcal/mol)	$\Delta G_{2,4}$ (kcal/mol)
$[(CAG):(GTC)]_4$	0.4	−4.0	−3.6
$[(CAG):(GTC)]_{10}$	9.8	−18.5	−8.7
$[(CAG):(GTC)]_{15}$	15.2	−31.5	−16.3
$[(CCG):GGC)]_4$	1.6	−6.2	−4.6

[a] As per Scheme 1, the free energy of duplex formation from individual hairpins ($\Delta G_{2,4}$) is simply the sum of the free energies of the two hairpin to single-strand transitions (ΔG_1) and the single-strand to duplex transition (ΔG_3) as reported in Table 2S.

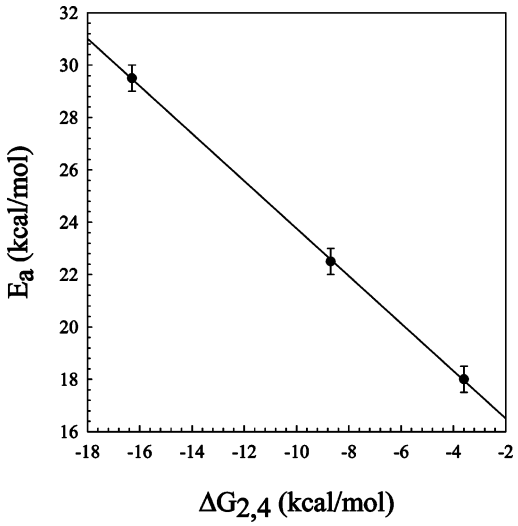
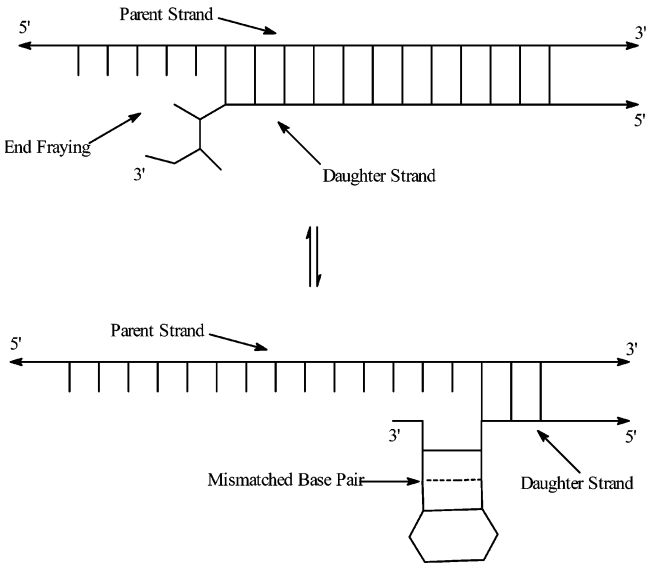


Figure 5. The correlation of the experimentally determined activation energy, E_a, to the calculated free energy of the hairpin to duplex transition, $\Delta G_{2,4}$, for the [(CAG):(GTC)] series.

favorable, as expected. It is interesting to note that, although the free energies of duplex formation become more favorable with increasing length, the corresponding activation energies become less favorable with increasing length.

As revealed in Figure 5, there is an interesting linear relationship between the free energy of duplex formation ($\Delta G_{2,4}$) and the activation energy for duplex formation (E_a). The implication, at least for the $[(CAG):(GTC)]_n$ series, is that the more stable the duplex, the slower is the rate of formation. In terms of the model proposed for gene expansion, the equilibrium between the fully duplexed parent:daughter species and the partially duplexed parent:hairpin species is dependent upon the relative stabilities of the two possibilities and the rate of the parent:hairpin to parent:daughter transition (Scheme 2). Thus,

Scheme 2. Equilibrium between the Fully Duplexed Parent:Daughter Species and the Partially Duplexed Parent:Hairpin Species



as previously shown, longer sequences of (CNG) triplets are more stable than shorter sequences,[5] and this work has demonstrated that longer hairpins are slower to reanneal with their complement. Together, these results are consistent with the conclusion of Gacy and McMurray.[4b] If gene expansion is partially due to hairpin formation on the newly synthesized daughter strand, as proposed by Wells,[2] the dependence on length on the lifetime of that hairpin can be explained in terms of thermodynamics and kinetics. What remains to be addressed is the mechanism of the parent:daugter to parent:hairpin transition, a current focus in this lab.

Acknowledgment. We thank Prof. George Turner for his critical reading of this manuscript. We also thank Merck, Inc. for their support of A.M.P. through their Doctoral Fellowship Program.

Supporting Information Available: A table of the line parameters for Figure 1 (Table 1S), a table of previously reported free energies of hairpin to single-strand transitions and duplex to single-strand transitions (Table 2S), and a figure of a representative absorbance decay for duplex formation (Figure 1S). This material is available free of charge via the Internet at http://pubs.acs.org.

JA043783N