The Consensus Sequence for Self-catalyzed Site-specific G Residue Depurination in DNA*

Olga Amosova, Alexander Smith, and Jacques R. Fresco

From the Department of Molecular Biology Princeton University, Princeton, New Jersey 08544

Background: Certain stem-loop-forming sequences self-catalyze site-specific DNA depurination of G residues.

Results: The catalytic intermediate is highly sequence-specific.

Conclusion: Like other catalytic mechanisms inherent in macromolecules, self-catalyzed DNA-depurination involves critical sequence and structural elements.

Significance: Because the resultant apurinic sites are subject to highly error-prone repair, knowledge of the sequence requirements enables location of potential spontaneous mutagenic sites within genes and other genomic elements.

Previously (1), we described a 14-residue DNA consensus sequence that, upon forming a stem-loop structure, self-catalyzes depurination of a specific G residue under physiological conditions in vitro some 4–5 orders of magnitude faster than the background “spontaneous” depurination rate of DNA in vivo (2). The self-depurinating single-strand element was originally identified within the coding strand of the human β-globin gene, with the depurinating G residue immediately upstream of the sickle cell anemia mutation site in codon 6. The catalytic intermediate was shown to be a stem-loop with a 5’-G(T/A)GG-3’ loop and T-A as the first base pair of the stem, followed by any sequence of several more base pairs sufficient to stabilize the structure. The 5’-G loop residue was shown to convert to an apurinic site susceptible to cleavage by either apurinic endonuclease or by spontaneous β-elimination at mildly alkaline pH. Such sites had previously been found to lead to substitution mutations or short deletions as a consequence of error-prone repair (3). In a preliminary search of the human genome, based upon the 14-residue self-depurinating sequence identified in the β-globin gene, at least 50,000 sites capable of forming such self-depurinating stem-loops were found, and when several variants corresponding to the consensus sequence were tested, they all showed the self-depurinating activity (1).

Given the possible biomedical and evolutionary consequences of this surprising self-catalytic reaction inherent in genomic DNA, we have sought to extend these initial observations to provide a more complete definition of the consensus sequence by extensively analyzing the consequences of varying the sequence of the stem-loop catalytic intermediate on the self-depurinating activity. This knowledge can provide a basis for elucidating the biological roles of the mechanism through an analysis of the frequency and distribution of potential self-depurinating sites in various genes and genomes, and it may also shed some light on the nature of the catalytic mechanism. Toward that end, we have examined the tolerance for replacement of each of the three G residues of the loop with canonical base or base analog residues, the tolerance for variation of the base pair at the base of the loop of the stem-loop catalytic intermediate, and the consequence of varying some features of the stem. As a result of this more precise definition of the consensus sequence, the number of potentially self-depurinating stem-loop-forming sequences in the human genome has been determined to be very much larger than originally estimated.

EXPERIMENTAL PROCEDURES

Deoxyoligonucleotides—These were purchased with and without base analogs from IDT (Coralville, IA), purified to homogeneity by denaturing PAGE (16% slab gels 25 × 45 cm in 8 M urea with Tris-borate buffer (90 mM Tris, 90 mM Boric acid, 0.5 M NaCl, pH 8.3)
Acid, 10 mM EDTA) at 1.5 kV for ~2.3 h), and eluted from gel slices by soaking overnight in 0.1 M sodium phosphate, pH 7.0. Final purification was accomplished by 100% acetonitrile elution from a C18 SEP-PAK reverse phase column (Waters), followed by spin evaporation. Oligomer concentration was then adjusted spectrophotometrically in Aldridge ACS reagent grade water.

3’-End Labeling—This was performed using terminal deoxynucleotidyl transferase (United States Biochemical Corp.) (5 units) on ~5 pmol of purified oligomer with ~5 pmol of [α-32P]dATP (4). The labeled oligomers were purified by denaturing PAGE prior to use.

Kinetics of Depurination—Oligomers were incubated at 22 °C in 10 mM sodium cacodylate, pH 5.0 (unless otherwise noted), and aliquots were frozen on dry ice at appropriate intervals. To reveal apurinic sites, aliquots were treated with 0.1 M piperidine for 30 min at 75 °C to induce base-catalyzed backbone cleavage (5, 6). Either a trace amount of end-labeled oligomer was mixed with an unlabeled stock of appropriate concentration prior to incubation or else the aliquots were labeled following the piperidine treatment (both methods yield the same depurination rates). The radioactivity of intact oligomer, of its 3’-end cleavage product, and of the products of nonspecific cleavage at other G residues was quantitated using a Molecular Dynamics store phosphorimaging device (American Biosciences), and these data were used for the kinetic analysis. Each experiment was performed at least three times, and the kinetic data were averaged.

RESULTS

Varying the Loop Residues—The stem-loop catalytic intermediate for self-catalyzed depurination of the 5’G residue of the loop in the wild-type β-globin gene is shown schematically in Fig. 1. Fig. 2 illustrates how the depurination kinetics was revealed by denaturing PAGE and the experimental data were quantitated to determine the depurination rates for different stem-loop sequences. Fig. 2A shows a typical gel fractionation after piperidine-induced cleavage of apurinic sites of self-depurinating oligomers (or the absence of such cleavage for non-depurinating ones), and Fig. 2B presents typical kinetic plots obtained from analysis of such gels. Data from those plots provide the basis for Table 1, which indicates how various substitutions of each of the loop residues and the residues of the first base pair with other canonical bases and base analogs of G affect the kinetics of self-depurination.
Consensus Sequence for G Residue Self-depurination in DNA

| First 5’ bp residue | Loop residues | First 3’ bp residue | Rate at pH 5.0 |
|---------------------|---------------|---------------------|----------------|
| T                   | G1            | A                   | 200 × 10⁻⁶ min⁻¹ |
| G                   | T2            | C                   | 60 × 10⁻⁶ min⁻¹ |
| A                   | G3            | T                   | <10⁻⁷ min⁻¹ |
| C                   | H             | G                   | <10⁻⁷ min⁻¹ |
| H                   | A             | 7-deaza-G           | <10⁻⁷ min⁻¹ |
| C                   | T             | 7-deaza-G           | <10⁻⁷ min⁻¹ |
| A                   | 7-deaza-G     | 7-deaza-G           | <10⁻⁷ min⁻¹ |
| 7-deaza-G           | 7-deaza-G     |                     | <10⁻⁷ min⁻¹ |

At loop position 1, the only other residue to undergo self-depurination at a rate close to that of dG is deoxynosine (Fig. 2B). The base in this case is hypoxanthine, which like G has a carbonyl substituent on C6, but unlike it, lacks the 2-amino substituent. Notably, as can be seen in Table 1, the other canonical purine residue, that with the base A, shows a definite trace tendency to self-depurinate. However, not unexpectedly, residues with 7-deaza-G (in which N7 is replaced by a carbon) and the two canonical pyrimidines C and T show no detectable tendency toward self-deglycosyl bond hydrolysis.

In the initial report (1), data were presented showing that depurination of the G residue at loop position 1 occurs only when the base of the second loop residue is a T, as in the sickle cell mutant, or an A, as in the wild-type β-globin gene. Substitution of a C-residue at that site results in complete absence of self-depurination at position 1, as does substitution of a G residue.

At loop position 3, replacement of the normal G residue with an I residue causes a small diminution in the self-depurination kinetics at position 1 (Fig. 2B), whereas replacement with residues of 7-deaza-G, -C, -T, or -A results in complete absence of self-depurination of loop residue G1.

At loop position 4, replacement of the G residue with the analog I residue also diminishes the kinetics of self-depurination of residue G1, but it does not eliminate it. This diminution is perhaps due to the well known poor stacking of the base hypoxanthine, in contrast to the strong stacking of G. This might lend flexibility to the loop by not stacking strongly with the first base pair of the stem, but only with 7-deaza-G and -T residues (dA and dC substitutions were not tested) is self-depurination at position 1 totally suppressed. In sum, sequence variation within the loop residues is quite restricted, with some sequence flexibility tolerated only at loop position 2.

Varying the Base Pair at the Base of the Loop—In the β-globin gene, the base pair at the base of the loop in the stem-loop catalytic intermediate for self-depurination is 5’-T:A-3’, and as can be seen in Table 1, with this base pair, the rate of self-depurination at position 1 is highest. Fig. 2 illustrates the effect of changing this first base pair of the stem. The consequences are quite striking: although replacement of that pair by 5’-G:C-3’ results in a mere 3-fold diminution in the rate, replacing it with either 5’-A:T-3’ or 5’-C:G-3’ (or any base pair mismatch; data not shown) totally suppresses self-depurination. A mismatch at the base of the loop likely permits a level of loop flexibility that removes any strain energy required for the depurination to occur. The difference between 5’-G:C-3’ and 5’-T:A-3’ versus 5’-C:G-3’ and 5’-A:T-3’ may well reside in the capacity of the similarly located carbonyls of T and G to hydrogen-bond to a water molecule so as to orient it toward the glycosyl bond of dG1 to be hydrolyzed, which lies just above it. Hence the differential base pair effects at the base of the loop.

Varying Stem Stability—To this point, all indications are that the lifetime of some fixed configuration of the loop impacts significantly on the probability of self-depurination to occur. The lifetime of such a strained loop configuration is likely to depend on the stability of the helically twisted stem, which in turn will depend on several well recognized variables, including helix length, i.e. the number of base pairs (7–10), base pair composition (11), and sequence (12), i.e. A:T, T:A, G:C, and C:G, noncomplementary base pair mismatches (13, 14), and the presence of extrahelical residues (13, 15). In fact, the data in Table 2 and the plot in Fig. 3 reveal a positive exponential correlation between the stability of these stem-loops, as indicated by their Tm values, and the rate of their self-depurination. Such a dependence is consistent with the two-state model of DNA melting, in which the proportion of time in stem-loop versus single-stranded conformation at any particular temperature should vary with the log of the difference between that temperature and Tm for the stem-loop (16).

Consistent with this model, we have found that a stem-loop with a short but G-C-rich stem without mismatches (D14p) displays almost the same depurination rate as an A-T-rich one, but with a 16-base pair-long (A:T)b stem (D(AT)b)16. Moreover, stem-loops with differences in stem sequence that have only a small effect on their Tm values (D18 and D18ch) display similar self-depurination rates. On the other hand, as seen in the
sequences in Table 2 and the PAGE analysis and plots in Fig. 4, replacement of the A-C mismatch in the stem of D18 (T_m of 43 °C, self-depurination rate of 2 \times 10^{-4} \text{ min}^{-1}) by a G-C base pair, as in D20p, leads not only to a very large increase of 26 °C in the T_m value, but also to an increase in the self-depurination rate so large (~40-fold) that it could only be estimated. Indeed, in this stem-loop, with its 7-base pair fully complementary stem with a G-C content of 57%, self-depurination was so rapid that in the hour it took at room temperature to purify it, the rate was too fast to permit more than an estimate that it is higher than 80 \times 10^{-4} \text{ min}^{-1}, and when a stem with a slightly different sequence of 8 base pairs with a G-C content of 62.5% (D22p) was prepared, its T_m rose another 4 °C, and the rate rose another 2–3-fold.

Effect of pH on the Rate of Self-catalyzed Depurination—The data presented above were all obtained at the standard pH 5 conditions to enable direct comparison with the previous observations (1). In these earlier experiments, a fragment of the human β-globin gene was used that contained an A-C mismatch in the stem of the catalytic intermediate. For this sequence and its various derivatives, it was shown that raising the pH to the physiological level of 7 significantly reduces the depurination rate. Such reduction is consistent with the well known fact that spontaneous DNA depurination is acid-catalyzed (19, 20) via protonation of both G and A residues, but it is also consistent with the stabilization of the A-C mismatched base pair at more acidic pH (21, 22). To determine whether the faster self-depurination rate at lower pH should be taken to mean that self-catalyzed DNA depurination is limited to in vitro conditions, the pH dependence was determined of the self-depurination rate for deoxyoligonucleotides with and without an A-C mismatch in the stem. As can be seen in the plots in Fig. 5, although the rate of self-depurination for the stem-loop with the A-C mismatch is reduced ~50-fold when the pH is raised from 5 to 7, it is only reduced ~4-fold when the...
mismatch is replaced by a G-C base pair, which results in a stem-loop with a fully paired and more stable stem. This difference is consistent with stabilization of the A-C mismatch at lower pH either by protonation of its A residue to form a wobble-type A⁺⁻C base pair (21) or of its C residue to form an A-C⁺⁻imino, enol complementary-type base pair (22). The comparison in Fig. 5 shows that only a modest fraction of the rate reduction at pH 7 is, in fact, due to the reduction in self-depurination per se. Apparently, the major effect at pH 7 (where the self-depurination rate greatly exceeds the background rate of spontaneous depurination) is due instead to the resulting destabilization of the A-C mismatch and the consequent drastic decrease in the stability of the stem, which, as was demonstrated by the data in Table 2 and the plot in Fig. 3, is a major determining factor in the self-catalyzed depurination rate. In vivo, however, it is very likely that intranuclear macromolecular crowding (23), which is known to elevate acidic pK values (because of the reduction in the proximity of the charged surfaces of the macromolecules and the consequent favorable base pairing that ensues) (24), both mitigates this destabilization and enables the required protonation of the purine residue that will undergo glycosyl bond hydrolysis. We have tested this rationalize by mimicking the macromolecular crowding with a well known crowding agent, polyethylene glycol 1000 at a concentration of 20% (conditions that we have found not to otherwise alter DNA helix stability). This resulted in ~10-fold increase in the in vitro self-depurination rate at pH 7 of the β-globin stem-loop consensus sequence (containing the A-C mismatch in the stem) to $8 \times 10^{-4}$ min⁻¹ (from $0.6 \times 10^{-6}$ min⁻¹, as shown in Fig. 5). This self-depurination rate is some 3 orders of magnitude faster than the background spontaneous depurination rate, which is high enough to justify including consensus sequences with a mismatch in the stem-forming sequence among the potentially self-depurinating sites in the genome.

DISCUSSION

The primary aims of this investigation were to define more completely the consensus sequence for site-specific self-catalyzed depurination of G residues within DNA sequences and then to evaluate the pH dependence of the mechanism. Aside from enabling much more accurate estimates of their presence in genomes, this information can be of value in identifying potential self-depurinating sites in genes of particular functional annotation and thereby provide a means of uncovering possible biological roles of the self-depurinating mechanism. The results presented show that the sequence variations examined in the loop and base pair at the base of the loop form a meaningful pattern in terms of the way they impact the self-depurination rate. Similarly, there is consistency between the stem sequence of the catalytic intermediate, the effect of pH variation, and its physical consequences on the stability of the stem, on the likely lifetime of the intermediate, and therefore on the rate of self-depurination. The consensus sequence would therefore appear to be suitably defined.

By varying stem parameters, it was found that self-depurination rates of deoxynucleotides capable of forming stem-loops with the appropriate loop sequence and perfectly paired stems can be $\geq 10^{-4}$ min⁻¹ at physiological pH (e.g. Fig. 5) or even higher when the stem-loops are more stable. Hence, self-depurination rates 4–5 orders of magnitude faster than the background spontaneous depurination rate (2) are quite possible.

The stem-loop structure essential for this mechanism can readily extrude from the DNA duplex (as part of a cruciform) under several circumstances: under the topological stress of DNA superhelicity (25); when the DNA strands separate during transcription or replication (26); and during ATP-dependent chromatin remodeling (27). Such extrusion has even been directly observed in genome sequences undergoing recurrent translocations in human chromosomes (28). All of these indications that cruciform extrusion can or does occur moves the observed phenomenon of self-catalyzed site-specific DNA depurination from a class of peculiar properties of some DNA sequences under unnatural conditions to one of potential biological relevance.

In fact, with the consensus sequence in hand, it has become possible to find strong indications of such relevance in the frequency and patterns of occurrence of sites in a genome with a potential to self-depurinate G residues. Such a mechanism may well have been expected to disappear over the course of evolution, given its capacity to damage the integrity of the genome as a consequence of error-prone repair of the resultant apurinic sites. Contrary to such expectation, a complete search of the human genome based upon the defined consensus sequence, which allows for a single base mispair or a single extrahelical residue (bulge) after the second base pair of the stem, both of which are thermodynamically plausible, reveals $\sim 1,000,000$ potential self-depurination sites for G residues. This number is some 5-fold higher than would be expected on a random basis, and this overrepresentation is even more prominent in many disease-associated genes, as well as some genes of common functional annotation that require sequence diversity in their contemporary function or else appear to have used the mechanism in the course of their evolution (29). Each such site has a potential for a substitution or short deletion mutation from error-prone repair of resulting apurinic sites. Although the extent to which potentially self-depurinating sequences in the human genome have been employed to create sequence diver-
sity is not presently apparent, the fact that significant examples have been uncovered emphasizes the value of knowing more fully the extent of sequence variation that retains the capacity to self-depurinate and so create potential apurinic mutagenic sites. In this connection, we have found many examples of coincidence of reported single-nucleotide polymorphisms with the consensus sequence described here for self-depurination of G residues. Prominent among such sites is the only one in the human β-globin gene; that site coincides with substitutions responsible for six different anemias and short deletions associated with two β-thalassemias (29).

The demonstration in this study that self-depurination can occur at a finite albeit low rate under essentially physiological conditions is worthy of note, particularly because of the implication that this mutation-related mechanism has likely played an evolutionary role. Although the low rate might be seen to indicate a mechanism of low significance, our view is that it is in fact appropriate for a mechanism likely to have been selected for creating sequence change in a biological context, in a manner analogous to the evolutionary sequence changes that arise from spontaneous mutations associated with the low error rates of highly accurate DNA synthesizing polymerases. Thus, both these mechanisms are geared to create genetic changes without creating cataclysmic genomic instability. Finally, we note that the tolerance for sequence variation in the consensus sequence for self-depurination of G residues that has been revealed and its affirmed acid catalysis provide insights whose further exploration should lead to a chemical mechanism for this unanticipated self-catalyzed reaction.

Acknowledgment—We are grateful to J. R. Alvarez-Dominguez for a careful reading of the manuscript.

REFERENCES
1. Amosova, O., Coulter, R., and Fresco, J. R. (2006) Proc. Natl. Acad. Sci. U.S.A. 103, 4392–4397
2. Lindahl, T., and Nyberg, B. (1972) Biochemistry 11, 3610–3618
3. Marnett, L. J., and Plastaras, J. P. (2001) Trends Genet. 17, 214–221
4. McEwan, N. R. (1997) Biotechnol. Techniques 11, 727–728
5. Sugiyama, H., Fujiwara, T., Ura, A., Tashiro, T., Yamamoto, K., Kawanishi, S., and Saito, I. (1994) Chem. Res. Toxicol. 7, 673–683
6. Lhomme, J., Constant, J. F., and Demeunynck, M. (1999) Biopolymers 52, 1465–83
7. Fresco, J. R., Alberts, B. M., and Doty, P. (1960) Nature 188, 98–101
8. Lipsett, M. N., Heppel, L. A., and Bradley, D. F. (1960) Biochim. Biophys. Acta 41, 175–177
9. Brahms, J., Maurizot, J. C., and Michelson, A. M. (1967) J. Mol. Biol. 25, 465–480
10. Blake, R. D., and Fresco, J. R. (1973) Biopolymers 12, 775–786
11. Marmur, J., and Doty, P. (1959) Nature 183, 1427–1429
12. Ornstein, R. L., and Fresco, J. R. (1983) Biopolymers 22, 1979–2000
13. Lommen, A. J., and Fresco, J. R. (1975) Prog. Nucleic Acid Res. Mol. Biol. 15, 185–218
14. Brown, T., Leonard, G. A., Booth, E. D., and Kneale, G. (1990) J. Mol. Biol. 212, 437–440
15. Fresco, J. R., and Alberts, B. M. (1960) Proc. Natl. Acad. Sci. U.S.A. 46, 311–321
16. Cantor, C. R., and Schimmel, P. R. (1980) Biophysical Chemistry: Part III. Behavior of Macromolecules, Chapters 22–3, W. H. Freeman, San Francisco, CA
17. Markham, N. R., and Zuker, M. (2005) Nucleic Acids Res. 33, W577–W581
18. Markham, N. R., and Zuker, M. (2008) in Bioinformatics, Structure, Function and Applications (Keith, J. M., ed) Vol. II, pp. 3–31, Humana Press, Totowa, NJ
19. Tamm, C., Hodes, M. E., and Chargaff, E. (1952) J. Biol. Chem. 195, 49–63
20. Suzuki, T., Ohsumi, S., and Makino, K. (1994) Nucleic Acids Res. 22, 4997–5003
21. Topal, M. D., and Fresco, J. R. (1976) Nature 263, 289–293
22. Purrello, R., Molina, M., Wang, Y., Smulevich, G., Fossella, J., Fresco, J. R., and Spiro, T. G. (1993) J. Am. Chem. Soc. 115, 760–767
23. Minton, A. P. (2001) J. Biol. Chem. 276, 10577–10580
24. Dill, K. A. (2002) Molecular Driving Forces: Statistical Thermodynamics in Chemistry & Biology, p. 418, Garland Science, London
25. Amosova, O., Kumar, V., Deusd, A., and Fresco, J. R. (2011) J. Biol. Chem. 286, 36322–36330
26. Dayn, A., Malkhosyan, S., Mirkin, S. M. (1992) 20, 5991–5997
27. Havas, K., Flaus, A., Phelan, M., Kingston, R., Wade, P. A., Lilley, D. M., and Owen-Hughes, T. (2000) Cell 103, 1133–1142
28. Kurahashi, H., Inagaki, H., Yamada, K., Ohye, T., Taniguchi, M., Emanuel, B. S., and Toda, T. (2004) J. Biol. Chem. 279, 35377–35383
29. Fresco, J., Amosova, O., Wei, P., Alvarez-Dominguez, J., Blumcher, D., and Torres, R. (2011) in Evolutionary Biology: Concepts, Biodiversity, Macroevolution and Genome Evolution (Puntarotti, P., ed) pp. 3–19, Springer-Verlag, Berlin