

Chapter 1

Site-Specific Self-Catalyzed DNA Depurination, the Basis of a Spontaneous Mutagenic Mechanism of Wide Evolutionary Significance

Jacques R. Fresco, Olga Amosova, Peter Wei, Juan R. Alvarez-Dominguez,
Damian Glumcher, and Rafael Torres

Abstract This chapter focuses on the nature of site-specific self-catalyzed DNA depurination as a spontaneous mechanism inherent in the chemical structure and dynamics of DNA that has contributed to evolutionary change. It describes the essential molecular features of the mechanism, the short consensus sequence elements that form the catalytic intermediate, the basics of the reactions that lead to the creation of apurinic sites, and the means by which those sites give rise to substitution and short deletion mutations. The consensus sequences are widely distributed in double-stranded genomes across the phyla at high frequency that increases up the phylogenetic tree. In the human genome, they constitute $>2 \times 10^6$ potential mutagenic sites, non-randomly scattered among very many genes, some containing multiple sites. Examples are presented of genes in which the mutations coincide with their self-depurination consensus sequences, the most striking being those in the β -globin gene that are responsible for six anemias and two β -thalassemias. Those of the olfactory receptor genes and the hypervariable regions of the immunoglobulin genes are shown to have utilized the mechanism to evolve their high degree of diversity and/or to develop their contemporaneous diversity for their present function.

1.1 Introduction

Spontaneous mutations are among the primary engines of evolutionary change. Until now, the major mode of their occurrence has been thought to be a consequence of errors in DNA replication, resulting in substitution and frameshift mutations. In reality, such mutational errors are not due to the enzymatic process having gone awry. Rather, substitution mutations are the consequences of intrinsic

J.R. Fresco • O. Amosova • P. Wei • J.R. Alvarez-Dominguez • D. Glumcher • R. Torres
Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA
e-mail: jrfresco@princeton.edu

the action of the widely occurring enzyme apurinic endonuclease (Korolev 2005) or by way of the well-known spontaneous β -elimination reaction (Lhomme et al. 1999; Sugiyama et al. 1994), the strand break site results in a frayed end accessible to exonuclease attack, particularly if the first base pair after the break is A•T, and so leads to a short deletion mutation.

In this report, some essential features of the underlying self-depurination mechanism are first described. We then proceed to issues of biological relevance, including indications that the mechanism has played a role in the evolution of some biological phenomena. Because the consensus sequence for self-depurination of G-residues was discovered several years ago, and that for A-residues only very recently, our data is much more extensive for the former. Nevertheless, it will become apparent that both mechanisms are very similar, and appear to have played comparable mutagenic roles in several biological processes, including those of molecular evolution.

1.2 Essential Features of DNA Self-Catalyzed Depurination

Self-catalyzed depurination is a remarkably site-specific reaction that does not involve the direct participation of either a protein enzyme or any multivalent cation or cofactor, and can occur under essentially physiological conditions (Amosova et al. 2006). As such, it represents the first natural deoxyribozyme activity discovered. It is mediated by formation of two very similar stem-loop-forming consensus sequences, which we have found to be present at rather high frequency in every double-stranded DNA genome searched, from the lowest to the highest form.

Figure 1.1 shows schematically the stem-loop structure of two sequences that are highly site-specific for self-catalyzed depurination of a G-residue (left) and an A-residue (right). As the figure indicates, the self-depurination mechanism removes the 5'G-residue of the loop in the former case, and the A-residue toward the 5' end in the latter one. In either case, the initial product of the catalytic event is an apurinic site in the loop sequence. This chain backbone site is thereby labilized, carrying a potential for intracellular backbone cleavage either by the enzyme apurinic endonuclease, or else as a result of its susceptibility to spontaneous backbone cleavage by a β -elimination reaction that can occur at slightly alkaline intracellular pH. Such an apurinic site, susceptible to error-prone repair, is potentially highly mutagenic and can give rise to a substitution or a short deletion. It is this resultant mutagenic potential that confers on self-depurination a role in evolution, all the more so because the self-catalytic depurination rate we have measured in vitro (Amosova et al. 2006) occurs 10^4 – $10^5\times$ faster than the background spontaneous depurination rate that has been estimated in vivo (Lindahl and Nyberg 1972).

DNA is typically double-stranded, whereas the catalytic intermediate for self-depurination is a single-stranded stem-loop. Hence, the self-depurination mechanism requires that the inverted repeat sequence harboring the self-depurinating loop

first extrude as a cruciform (half of which contains the single-stranded self-depurinating stem-loop), and that the cruciform have a sufficient lifetime for the reaction to occur. That such cruciform extrusion can take place has been demonstrated previously (Alvarez et al. 2002; Inagaki et al. 2009; Kim et al. 1998; Shlyakhtenko et al. 1998). We have recently performed experiments under physiological conditions *in vitro* with stem-loop-forming sequences for G-residue self-depurination embedded in supercoiled plasmids, in which such extrusion has been directly shown to be crucial for the self-catalytic depurination (Amosova et al. 2011b).

As Fig. 1.1 indicates, the essential features of the stem-loops are very similar, i.e., both have tetra-loops with different highly specific sequences: 5'G-A/T-G-G for depurination of the extreme 5' G-residue and 5'G-A-G-A for depurination of the second residue in from the 5' end, which is an A. Interestingly, both can form a homopurine base pair within the loop, G⁺•G and A⁺•A, in which the residue to be depurinated is protonated at N3 of the base (Lavelle and Fresco, in preparation). It is this base pair formation that likely explains why acid-catalyzed depurination can actually occur in the neutral pH environment of most cells.

The G-residue self-depurination activity exhibits very limited tolerance for loop sequence variation (Amosova et al. 2011a). For example, the three G-residues in the loop are replaceable only by hypoxanthine, a closely related purine analogue, with only modest reduction in the activity; and the A-residue in the G-self-depurinating sequence is replaceable only by a T-residue, in this case with activity enhancement. In contrast, there is total tolerance for variation in the complementary base pairs of the helical stem, except for the first one at the base of the loop. Even a single base pair mismatch or a single extrahelical base elsewhere in the stem can be tolerated. Apparently, the role of the stem is to stabilize and maintain the loop in some strained configuration favorable to glycosyl bond cleavage. Thus, the more stable the stem, as affected by length (Blake and Fresco 1973; Brahms et al. 1967), G•C content (Marmur and Doty 1959), base pair sequence (Ornstein and Fresco 1983), base pair mismatches (Lomant and Fresco 1973), and the presence of extrahelical bases (Lomant and Fresco 1973), the faster the rate of self-depurination. The nature of the first base pair is somewhat restricted, possibly because it orients the water molecule involved in the hydrolysis of the glycosyl bond of the residue to be depurinated.

1.3 Biological Relevance of the Self-Depurination Mechanism

The self-depurination reaction was discovered in the course of working with 29-residue long complementary deoxyoligonucleotide strands of the human β -globin gene that contained the sickle cell anemia mutation site. Whereas the duplex formed from those strands, as well as the noncoding strand by itself, were characteristically stable, the coding strand was not. Rather, under solvent, pH, and temperature that mimic physiological conditions, it rapidly self-fragmented not

randomly, but in a unique way. This fragmentation was found to arise from backbone cleavage that was ultimately traced to spontaneous β -elimination at the apurinic site caused by self-catalyzed depurination mediated by stem-loop formation (Amosova et al. 2006). The occurrence of such a reaction in a strand segment of a significant human gene, immediately upstream of the sickle cell mutation site, provided the impetus for trying to understand the significance of what was obviously a DNA self-catalyzed reaction.

With the finding that the stem base-paired sequence is tolerant of great variation, it was decided to determine the number of potential stem-loops for self-depurination of G-residues in the human genome, and their distribution across the phyla. The numbers proved to be surprisingly large, and indicative of substantial overrepresentation relative to random expectation, which was not the case for stem-loops of similar size with non-self-depurinating loop sequences. The overrepresentation was therefore taken to be indicative of some important bio-functionality of the stem-loops for self-depurination. The whole-genome search for self-depurinating stem-loops was complemented first by identifying all human genes containing those sequences, including the loci of their occurrence, i.e., exons, introns, control elements, untranslated regions, intergenic regions, etc. This was followed by analysis of G-residue consensus sequence occurrence within more than 100 individual genes and their degree of overrepresentation relative to random expectation (see Table 1.3 as an example). These searches, together with the uncovering of the stem-loop consensus sequence for self-depurination of A-residues by way of its overrepresentation in the human genome, led us to the indications of the role of self-depurinating sequences in evolution; and it is from this vantage point that we present the findings that follow.

1.4 Self-Depurination in the Human β -Globin Gene

This gene, which has 148 codons (in its message), contains no A-residue self-depurination site, and but one site for G-residue self-depurination, of which the first three residues of the loop correspond to codon 6. It is the second residue of the loop that is the site of the sickle cell anemia mutation. Figure 1.2 shows a plot of the number of independent variations (i.e., those in different haplotypes) per codon of this gene. The plot is based upon data obtained from two databases, HB Var (<http://globin.cse.psu.edu>), which includes all human hemoglobin variants and β -thalassaemia mutations reported in the literature over more than half a century, and the recently assembled Human Gene Mutation Database (HGMD) (<http://www.hgmd.cf.ac.uk>). As such, the plot represents a compendium of all unique β -globin alleles revealed in most of the human populations in the world. It is striking that the most prominent mutation site in the plot in Fig. 1.2, that at codon 6, corresponds to three of the four loop residues of that single self-depurinating consensus sequence in this gene. In this glutamic acid codon, residue #1, the self-depurinating G-residue, and residue #2, the site of the sickle cell mutation, are the sites of readily detectable

Evolutionary Biology – Concepts, Biodiversity,
Macroevolution and Genome Evolution

Pontarotti, P. (Ed.)

2011, XV, 345 p., Hardcover

ISBN: 978-3-642-20762-4