Guanosine Distribution and Oxidation Resistance in Eight Eukaryotic Genomes

Keith A. Friedman* and Adam Heller

*Contribution from the Department of Chemical Engineering and the Texas Materials Institute,
University of Texas at Austin, Austin, Texas 78712-0231*

Received August 29, 2003; E-mail: kafriedman@mail.utexas.edu

Abstract: Reactive oxygen species that attack DNA are continuously generated in living cells. Both the guanosine (G) mole fraction and its distribution should affect the stability of genomes and their parts to oxidation. At a lesser G content, genomes should be more oxidation resistant or "ennobled". Oxidant scavenging by G's in nonessential parts of introns and intergenic domains should decrease G oxidation in the essential exons. To determine whether genomes are indeed ennobled and whether oxidant-scavenging domains exist in genomes, the relative rates of guanosine oxidation in average exons, introns, and intergenic domains were estimated. Comparison among genomes indicated that average exons are ennobled in the genomes of *Caenorhabditis* (worm), *Arabidopsis* (plant), *Saccharomyces* (yeast), *Schizosaccharomyces* (yeast), and *Plasmodium* (malaria parasite), and that average introns and intergenic domains are ennobled in these genomes and in the genome of *Drosophila* (fly). The exon oxidation rates estimated for these genomes were less than the rate for the hypothetical "standard" genome, with a 0.25 mole fraction of uniformly distributed G. For *Plasmodium* the rate was half of that estimated for the standard genome. Average exons were not ennobled in the human or fly genomes; their G distributions were comparable to that in the standard genome. Instead, their exons were situated between introns and intergenic domains that could protect them by oxidant scavenging, the G's of their introns and intergenic domains outnumbering those of their exons 50-fold in humans and 4-fold in flies. The G distribution in the *Encephalitozoon* (parasite) genome was not protective relative to that of the standard genome.

Introduction

Although the oxidation potentials of the bases A, T, G, and C are comparable to those of the noble metals Pt, Au, and Pd,[1,2] $\sim 10^4$ bases of the genome in each human cell are oxidized daily.[3,4] At steady state, the human genome has between 10^4 and 10^5 oxidative lesions,[5] including base and sugar lesions, strand breaks, and protein cross-links.[6] About 5% of oxidative lesions[7] or ~ 10 in 10^6 bases[8,9] are 8-oxo-7,8-dihydro-2′-deoxyguanosine, a mutagenic lesion[10] that is a common product of guanosine oxidation.[11] The several mechanisms for its repair in mammalian cells strongly suggest that it presents a genuine threat to genomic integrity.[6] These lesions occur and persist despite the enzymes, antioxidants, and multiple repair systems that actively protect the genome and despite the passive protection afforded by membranes and chromatin and by DNA's high oxidation potential.

The strongest oxidizers react with massive release of free energy and oxidize bases indiscriminately, but some strong oxidizers attack guanine bases selectively. G selectivity has been shown in vitro with genomic sequences[12–14] and in vivo with metal–H_2O_2 oxidants[15–17] and with mitochondria.[18] Some strong oxidizers (benzoyloxyl radicals, sensitized photooxidizers, etc.) are selective for G's in GG and GGG sequences, while others (metal-catalyzed H_2O_2, $^1O_2$, etc.) are not.[11,13] GG and GGG selectivity has been shown in vitro with genomic[12–14] and synthetic[19,20] sequences, with nucleosome core particles,[21] and

(1) Faraggi, M.; Broitman, F.; Trent, J. B.et al. *J. Phys. Chem.* **1996**, *100*, 14751–14761.
(2) Steenken, S.; Jovanovic, S. V. *J. Am. Chem. Soc.* **1997**, *119*, 617–618.
(3) Helbock, H. J.; Beckman, K. B.; Shigenaga, M. K.et al. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 288–293.
(4) Setlow, R. B. *Mutat. Res.* **2001**, *477*, 1–6.
(5) Beckman, K. B.; Ames, B. N. *J. Biol. Chem.* **1997**, *272*, 19633–19636.
(6) Cooke, M. S.; Evans, M. D.; Dizdaroglu, M.; et al. *FASEB J.* **2003**, *17*, 1195–1214.
(7) Dizdaroglu, M. *Mutat. Res.* **1992**, *275*, 331–342.
(8) Dizdaroglu, M.; Jaruga, P.; Birincioglu, M.; et al. *Free Radical Biol. Med.* **2002**, *32*, 1102–1115.
(9) Guetens, G.; De Boeck, G.; Highley, M.; et al. *Crit. Rev. Clin. Lab. Sci.* **2002**, *39*, 331–457.
(10) Shibutani, S.; Takeshita, M.; Grollman, A. P. *Nature* **1991**, *349*, 431–434.
(11) Burrows, C. J.; Muller, J. G. *Chem. Rev.* **1998**, *98*, 1109–1151.

(12) Kawanishi, S.; Oikawa, S.; Murata, M.; et al. *Biochemistry* **1999**, *38*, 16733–16739.
(13) Kawanishi, S.; Hiraku, Y.; Oikawa, S. *Mutat. Res.* **2001**, *488*, 65–76.
(14) Rodriguez, H.; Valentine, M. R.; Holmquist, G. P.; et al. *Biochemistry* **1999**, *38*, 16578–16588.
(15) Henle, E. S.; Luo, Y.; Gassmann, W.; et al. *J. Biol. Chem.* **1996**, *271*, 21177–21186.
(16) Henle, E. S.; Linn, S. *J. Biol. Chem.* **1997**, *272*, 19095–19098.
(17) Kawanishi, S.; Hiraku, Y.; Murata, M.; et al. *Free Radical Biol. Med.* **2002**, *32*, 822–832.
(18) Driggers, W. J.; Holmquist, G. P.; LeDoux, S. P.; et al. *Nucleic Acids Res.* **1997**, *25*, 4362–4369.
(19) Saito, I.; Nakamura, T.; Nakatani, K.; et al. *J. Am. Chem. Soc.* **1998**, *120*, 12686–12687.
(20) Hickerson, R. P.; Prat, F.; Muller, J. G.; et al. *J. Am. Chem. Soc.* **1999**, *121*, 9423–9428.
(21) Nunez, M. E.; Noyes, K. T.; Barton, J. K. *Chem. Biol.* **2002**, *9*, 403–415.

with isolated nuclei,[22] but it has not been shown in vivo and would vary among the various oxidants in cells whose concentrations would vary among organisms.

The impact of guanosine oxidation should depend on the genomic distribution of G's. Lower G percentages should lower the guanosine oxidation rate in genomes and genomic domains such as protein-coding exons, making them more noble.[23] G's in the nonessential portions of introns and intergenic domains (IGDs) should scavenge selective oxidants, reducing their concentrations and decreasing their attack on G's in exons.

Genomic G should be a significant scavenger of selective oxidants in the nucleus. Individual G's are unlikely to be overwhelmed by oxidants, because their oxidation frequency (years per exon) is much lower than their repair frequency (minutes to days per lesion).[24] The concentration of G in the nucleus is comparable to that of antioxidants: ~60 mM genomic G in humans and ~2 mM in *Arabidopsis* (plant)[25] vs ~1 to ~10 mM glutathione in mammalian cytoplasm[26] and plant chloroplasts.[27] However, freely diffusing antioxidants are more reactive than genomic guanosine,[28,29] which is buried in the double helix and shielded by histone proteins in nucleosomes.[21,30] Global scavenging by nonexon G's should reduce oxidant concentrations and the global rate of exon oxidation. In addition, local scavengers, e.g., G's in proximal introns and/or IGDs, could scavenge a hole or an oxidant near an exon and its splice sites.

Scavenging could be particularly effective when sacrificially oxidizable G's are near exons, because in vitro experiments have shown that oxidant-injected holes selectively react with remote GG and GGG,[21,22,31–34] because duplex DNA supports hole (electron vacancy) transport across 50–100 base pairs.[35–41] Barton[22,41–43] and co-workers have investigated this chemistry for more than a decade, and wrote the following: "However, one could consider that segments throughout the genome may encode 'sinks' for damage, and that other segments could serve as buffers as a result of local sequence-dependent or protein-dependent structural deformations to protect critical regions." Giese,[37] Kawanishi,[44] Thorp,[45] Heller,[23] and their co-workers have considered both sacrificial G anodes and scavenger G's, and Kanvah and Schuster[46] have shown that disulfides on molecules intercalated in DNA can serve as sacrificial anodes.

In this study, the G distributions were analyzed in eight eukaryotic genomes: *Homo sapiens* (*Hsa*, human), *Drosophila melanogaster* (*Dme*, fruit fly), *Caenorhabditis elegans* (*Cel*, nematode worm), *Arabidopsis thaliana* (*Ath*, flowering plant), *Saccharomyces cerevisiae* (*Sce*, budding yeast), *Schizosaccharomyces pombe* (*Spo*, fission yeast), *Plasmodium falciparum* (*Pfa*, malaria parasite), and *Encephalitozoon cuniculi* (*Ecu*, intracellular parasite). The relative rates of guanosine oxidation in average exons, introns, and IGDs were estimated. Comparison among genomes indicated ennoblement in the average exons of *Cel*, *Ath*, *Sce*, *Spo*, and *Pfa*, and in the average introns and IGDs of these genomes and the *Dme* genome. Ennoblement was not found in the average exons, introns, or IGDs of the human genome or in the average exons of *Dme*. Human and *Dme* exons were situated in the oxidant-scavenging environment of large introns and IGDs whose G's outnumbered those in exons 50-fold in *Hsa* and 4-fold in *Dme*. The G distribution in the *Ecu* genome did not appear protective relative to that in the standard genome.

Computational Methods

Determination of G Distributions from Genome Data. Genome sequences from GenBank[47] for *Hsa* (February 2002 release),[48,49] *Dme* (October 2000)[50] *Cel* (December 2001),[51] *Ath* (January 2002),[52] *Sce* (March 2002),[53] *Spo* (March 2002),[54] *Ecu* (March 2002),[55] and *Pfa* chromosomes[56] 2 (November 1998)[57] and 3 (April 1999)[58] were analyzed. The gene annotations were manipulated to merge overlapping or duplicated exons (protein-coding sequences or CDS).

Each G was identified as part of an exon, intron, or IGD if it was in a CDS, between exons, or between genes, respectively. The numbers of exons ranged from ~10^3 in *Ecu* to ~10^5 in *Hsa*. If an intron or IGD nucleotide was ≥6 bp and ≤105 bp from the boundaries of its segment, it was further identified as part of a flank. Flanks were defined as ≤100 bp long segments at the ends of introns or IGDs, i.e., flanking exons. The five nucleotides at the ends of introns and IGDs were omitted from flanks, because they are involved in exon/intron splicing. Weighted

(22) Nunez, M. E.; Holmquist, G. P.; Barton, J. K. *Biochemistry* **2001**, *40*, 12465−12471.
(23) Heller, A. *Faraday Discuss.* **2000**, *116*, 1−13.
(24) The 2 × 3 × 10⁹ bp in each human cell divided by 160 bp in an average-length exon and divided by ~10⁴ oxidations per day gives ~4000 days between attacks on an average exon. 8-Oxo-7,8-dihydro-2′-deoxyguanosine has a half-life of 11 min in rat liver.⁶⁶ About 10⁴ 8-oxo-7,8-dihydro-2′-deoxyguanosine bases at steady state divided by ~10⁴ oxidations per day gives ~1 day for repair, because the oxidation rate equals the repair rate at steady state.
(25) ΣG concentrations in a ~5 μm diameter nucleus were estimated from 2 × 3 × 10⁹ bp at 20% ΣG in the *Hsa* genome and 2 × 1 × 10⁸ bp at 20% ΣG in the *Ath* genome.
(26) Dringen, R. *Prog. Neurobiol.* **2000**, *62*, 649−671.
(27) Noctor, G.; Arisi, A.-C. M.; Jouanin, L.; et al. *J. Exp. Bot.* **1998**, *49*, 623−647.
(28) Maki, H.; Sekiguchi, M. *Nature* **1992**, *355*, 273−275.
(29) Hogan, M. E.; Rooney, T. F.; Austin, R. H. *Nature* **1987**, *328*, 554−557.
(30) Smerdon, M. J.; Thoma, F. *Contemp. Cancer Res.* **1942**, *2*, 199−222.
(31) Boone, E.; Schuster, G. B. *Nucleic Acids Res.* **2002**, *30*, 830−837.
(32) Meggers, E.; Michel-Beyerle, M. E.; Giese, B. *J. Am. Chem. Soc.* **1998**, *120*, 12950−12955.
(33) O'Neill, P.; Parker, A. W.; Plumb, M. A.; et al. *J. Phys. Chem. B* **2001**, *105*, 5283−5290.
(34) Sanii, L.; Schuster, G. B. *J. Am. Chem. Soc.* **2000**, *122*, 11545−11546.
(35) Lewis, F. D.; Liu, X.; Liu, J.; et al. *J. Am. Chem. Soc.* **2000**, *122*, 12037−12038.
(36) Bixon, M.; Giese, B.; Wessely, S.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11713−11716.
(37) Giese, B. *Annu. Rev. Biochem.* **2002**, *71*, 51−70.
(38) Nunez, M. E.; Hall, D. B.; Barton, J. K. *Chem. Biol.* **1999**, *6*, 85−97.
(39) Sistare, M. F.; Codden, S. J.; Heimlich, G.; et al. *J. Am. Chem. Soc.* **2000**, *122*, 4742−4749.
(40) Treadway, C. R.; Hill, M. G.; Barton, J. K. *Chem. Phys.* **2002**, *281*, 409−428.
(41) Boon, E. M.; Barton, J. K. *Curr. Opin. Struct. Biol.* **2002**, *12*, 320−329.
(42) Murphy, C. J.; Arkin, M. R.; Jenkins, Y.; et al. *Science* **1993**, *262*, 1025−1029.
(43) Hall, D. B.; Holmlin, R. E.; Barton, J. K. *Nature* **1996**, *382*, 731−735.
(44) Oikawa, S.; Tada-Oikawa, S.; Kawanishi, S. *Biochemistry* **2001**, *40*, 4763−4768.
(45) Szalai, V. A.; Singer, M. J.; Thorp, H. H. *J. Am. Chem. Soc.* **2002**, *124*, 1625−1631.
(46) Kanvah, S.; Schuster, G. B. *J. Am. Chem. Soc.* **2002**, *124*, 11286−11287.
(47) Benson, D. A.; Karsch-Mizrachi, I.; Lipman, D. J.; et al. *Nucleic Acids Res.* **2002**, *30*, 17−20.
(48) International Human Genome Sequencing Consortium. *Nature* **2001**, *409*, 860−921.
(49) Venter, J. C.; Adams, M. D.; Myers, E. W.; et al. *Science* **2001**, *291*, 1304−1351.
(50) Adams, M. D.; Celniker, S. E.; Holt, R. A.; et al. *Science* **2000**, *287*, 2185−2195.
(51) C. elegans Sequencing Consortium. *Science* **1998**, *282*, 2012−2018.
(52) Arabidopsis Genome Initiative. *Nature* **2000**, *408*, 796−815.
(53) Goffeau, A.; Barrell, B. G.; Bussey, H.; et al. *Science* **1996**, *274*, 546, 563−567.
(54) Wood, V.; Gwilliam, R.; Rajandream, M. A.; et al. *Nature* **2002**, *415*, 871−880.
(55) Katinka, M. D.; Duprat, S.; Cornillot, E.; et al. *Nature* **2001**, *414*, 450−453.
(56) Chromosomes 2 and 3 constitute only 7% of the *Pfa* genome, but their ΣG + ΣC mole percentage (20%) is typical of the whole genome (18%) (data in refs 57 and 58).
(57) Gardner, M. J.; Tettelin, H.; Carucci, D. J.; et al. *Science* **1998**, *282*, 1126−1132.
(58) Bowman, S.; Lawson, D.; Basham, D.; et al. *Nature* **1999**, *400*, 532−538.

Table 1. Estimated Relative Rates[a] of G Oxidation in the Average Exons, Introns, IGDs and Weighted Average Flanks of Eight Eukaryotic Genomes[b] and the Standard[c] Genome, Expressed as Percentages of the Rate in the Corresponding Domain of the Standard Genome, E.g., (*Hsa* Exons)/(*Std* Exons)

	Hsa	Dme	Cel	Ath	Sce	Spo	Ecu	Pfa	Std
Calculated with All G's Oxidized Equally (Eq 1)									
Exons	102	105	84	87	80	79	95	51	100
Introns	91	73	58	65	67	58	d	26	100
IGD	88	76	65	65	67	62	88	28	100
Flanks	92	67	50	60	63	54	71	26	100
Calculated with G's in GG and GGG Oxidized More Readily (Eq 3)									
Exons	108	103	75	81	74	71	94	44	100
Introns	99	66	49	54	57	45	d	20	100
IGD	93	70	58	56	59	51	91	23	100
Flanks	102	62	42	50	54	44	73	21	100

[a] These rates were normalized by the numbers of bases, so they do not indicate the relative amounts of oxidation in the domains. [b] *Hsa*, *Homo sapiens* (human); *Dme*, *Drosophila melanogaster* (fruit fly); *Cel*, *Caenorhabditis elegans* (nematode worm); *Ath*, *Arabidopsis thaliana* (flowering plant); *Sce*, *Saccharomyces cerevisiae* (bakers yeast); *Spo*, *Schizosaccharomyces pombe* (budding yeast); *Ecu*, *Encephalitozoon cuniculi* (intracellular parasite); *Pfa*, *Plasmodium falciparum* (malaria parasite); *Std*, standard.[c] All domains of the standard genome had 25% ΣG and G's allocated per probability (see the Supporting Information). [d] *Ecu* introns were omitted because there were fewer than 100.

average flanks were the averages of intron flanks weighted with the mean number of introns per gene and IGD flank weighted with one. When an intron or (rarely) IGD was <112 bp long, its flanks were identified as the domain ≥6 bp from both of its ends.

Because G's in GG and GGG sequences are more easily oxidized in vitro, G's were subtyped as **G**, **GG**−**GG** or **GGG**, where **G** was the lone G in HGH; **GG** was the left or 5′-G in GGH, **GG** was the right or 3′-G in HGG, **GGG** was the central G in GGG, H was any base but G, ΣG was total G, and **GG**−**GG** was **GG** or **GG**. (A bold **G** refers to these subtypes, whereas a plain G refers to guanosine generally.) For each segment, the ΣG, **G**, **GG**−**GG**, or **GGG** mole percentages in the segment overall and in its flanks were calculated. Each segment was weighted equally, regardless of length, to calculate mean mole percentages. These segment-weighted averages described average exons, introns, and IGDs, not average nucleotides. The mean mole percentages of each nucleotide type vs distance from the 3′ or 5′ segment boundaries were plotted.

Estimation of G Oxidation Rates. Approximating DNA as a simple material in aqueous solution and neglecting the differences in its environment in different organisms, a domain's total G oxidation rate was set proportional ($\propto$) to its total G concentration (eq 1). In this equation, the concentration in the nucleus was replaced by the mole percentage in the genome, to which it was proportional. Evaluating this equation using the mean mole percentages in Table S1 in the Supporting Information and dividing the rates in the eight actual genomes by the rate in the corresponding domain of the "standard" genome gave the relative rates of G oxidation in average exons, introns, and IGDs (Table 1). The standard genome was a hypothetical construct in which A, T, C, and G were equimolar and randomly distributed. It facilitated comparison among eukaryotic genomes, but was not itself a basis for valuation.

$$\Sigma\text{G oxidation rate} \propto [\Sigma\text{G}] \qquad (1)$$

domain ΣG oxidation amount $\propto$

$$(\Sigma\text{G oxidation rate}) \times (\text{domain/genome}) \quad (2)$$

$$\Sigma\text{G oxidation rate} \propto [\textbf{G}] + [\textbf{GG}] + 3[\textbf{GG}] + 5[\textbf{GGG}] \quad (3)$$

The relative rates in Table 1 do not indicate the relative amounts of G oxidation or "damage" in average exons, introns, or IGDs of the genomes, because the rates were normalized by the numbers of bases in the domains. Evaluating eq 1 and weighting the resulting rate by

Table 2. Estimated[a] Relative Amounts of G Oxidation or Damage to Exons, Introns, IGDs, and Weighted Average Flanks of Eight Eukaryotic Genomes, Expressed as Percentages of the Damage to the Entire Genome, E.g., (*Hsa* Exons)/(*Hsa*)

	Hsa	Dme	Cel	Ath	Sce	Spo	Ecu	Pfa
Exons	2	22	31	35	76	61	87	63
Introns	48	13	23	14	<1	3		2
IGDs	50	64	46	51	24	37	13	35
Flanks	2	7	17	17	8	8	11	5

[a] Calculated with all G's oxidized equally (eq 1). Calculations with G's in GG and GGG oxidized more readily (eq 3) gave very similar results (Table S3).

the fraction of the genome in that domain gave a value proportional to the amount of G oxidation (eq 2). Evaluating this equation using the data in Tables S1 and S2 and dividing the values by the sum of values in the corresponding genome gave the relative damage to guanosine in exons, introns, and IGDs, expressed as percentages of the damage to the entire genome (Table 2).

To assess the effects of putative differences in guanosine oxidation rates depending on neighbors, eq 1 was refined into eq 3 by expanding [ΣG] to [**G**] + [**GG**] + [**GG**] + [**GGG**] and multiplying their mean mole percentages by conservative estimates of their relative oxidation rates (see the Supporting Information). Evaluating this equation gave the results in Tables 1 and S3.

Results and Discussion

Table 1 lists the calculated relative rates of G oxidation in average exons, introns, IGDs, and flanks of the eight eukaryotic genomes, expressed as percentages of the rates in the standard genome. These numbers are indicative, but not quantitative, because of the approximations in their estimation. The ΣG mole percentage averaged 25 ± 1% in *Hsa*, *Dme*, and *Ecu* exons, ≤22% in *Cel*, *Ath*, *Sce*, *Spo*, and *Pfa* exons, and ≤20% in the introns and IGDs of all of the model genomes except *Hsa* (Table S1 in the Supporting Information). Hence, the calculated relative rates were lower in all average domains of *Cel*, *Ath*, *Sce*, *Spo*, and *Pfa* and in average introns and IGDs of *Dme* than in the corresponding domains of the standard genome. Characterizing domains with higher and lower G oxidation rates as "hotter" and "colder", *Hsa* was hotter than other eukaryotic genomes, and its exons were hotter than its other domains. Adding GG and GGG selectivity to the calculations made *Hsa* and *Ecu* exons, introns, and IGDs and *Dme* exons hotter and all other domains colder.

Figure 1 shows the following tendency: when the exon oxidation rate was higher (hotter), the exon fraction of genomic DNA was smaller, and, *vice versa*, when the exon oxidation rate was lower, the fraction of DNA in exons was larger. This was true for six genomes, but not for *Ecu* and *Sce*. Thus, when exons were hotter, they resided in a larger pool of oxidant scavenging introns and intergenic domains. When they were colder, the scavenger pool was smaller, and less benefit would have accrued of such a pool. This tendency was stronger with GG and GGG selectivity in the calculations.

Table 2 shows the calculated relative amounts of G oxidation or 'damage' to the exons, introns, IGDs, and flanks of the eight genomes, expressed as percentages of the damage to the entire genome. The ratios of the total number of G's in introns and IGDs to their total number in exons were 56 in *Hsa*, 1.9−3.5 in *Dme*, *Cel*, and *Ath*, and 0.2−0.6 in *Sce*, *Spo*, *Ecu*, and *Pfa* (Tables S1 and S2). Hence, about 50 times more G damage was predicted in *Hsa* introns and IGDs than in *Hsa* exons,
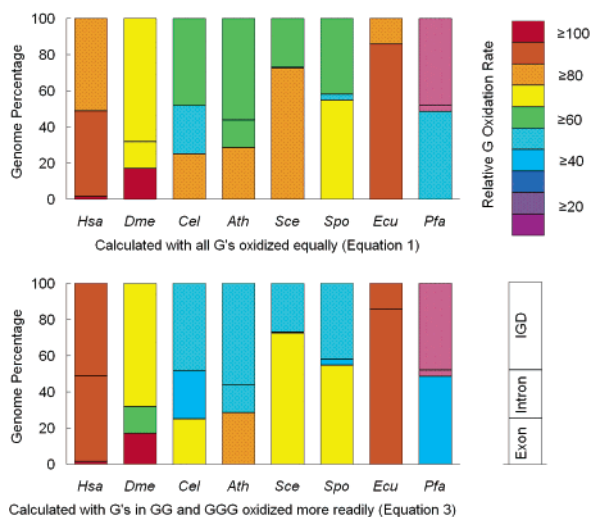
Figure 1. Fractions of the eight eukaryotic genomes in "hot or cold" exons, introns, or intergenic domains. The estimated relative rates of guanosine oxidation (Table 1) in average exons, introns, and IGDs are indicated by the colors of the bars. The percentages of the genomes that are exons, introns, and IGDs are represented by the bar heights.
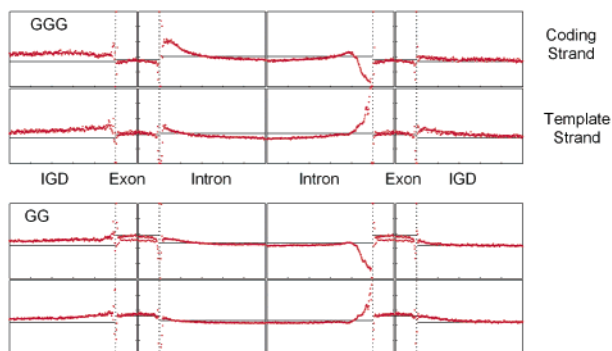

Figure 2. Variation of the mean mole percentages of GGG and GG with distance from exon, intron, or IGD boundaries (red dots) and the overall mean percentages of GGG and GG (black lines) in *Hsa*. Introns and IGDs longer than 100 bp are shown. The vertical full scale is 5% for GGG and 10% for GG. The horizontal full scale is 300 bp on all charts.

consistent with introns and IGDs functioning as oxidant sinks for the human genome. For *Dme*, *Cel*, and *Ath*, the calculated damage to introns and IGDs was 4−2 times greater than that to exons.

In *Hsa*, exons and weighted average flanks had similar G damage predicted (Table 2). The scavenging potential of *Hsa* flanks was enhanced by their elevated GG and GGG levels near exon boundaries (Figure 2). These elevations can have biological functions; e.g., GGG sequences in the 5′ flank of vertebrate

introns enhance splicing, particularly of small exons and introns.[59,60] Nevertheless, given a conductive path, G's in flanks could act as sacrificial anodes, drawing holes out of neighboring exons.[23,61] However, charge transport can be obstructed by DNA sequences and structures,[38,40,62] so the putative conduction path needs to be evaluated in each case.

The finding that genomic G distributions were consistent with reducing G oxidation rates in protein-coding exons by ennoblement or by oxidant scavenging is qualitative not quantitative and should not be overgeneralized. The rate estimates only indicated directions and ordering, because G oxidation rates differ between laboratory and living systems, and oxidant concentrations differ between organisms. G minimization effects probably were much greater and much less in some segments than they were on average, because individual segments varied widely from the averages. The percentage standard deviations of mole fractions, which indicated variation not imprecision, were ∼30% to ∼300%. This study assessed the affect of G levels, not the evolution of these levels. G levels can be related to selection and mutational pressures unrelated to oxidation.[63−65] In this study, essential domains were restricted to exons, not because they were the only essential domains, but because they were the only ones that were extensively annotated in the genomes. Assessing the importance of guanosine-based protection against oxidation relative to that of other defenses was beyond the scope of this study. Different organisms can and do specialize in different defenses.

Supporting Information Available: GG and GGG selectivity of G oxidation, methods used for probability predictions of guanosine distributions, and Tables S1−S3 (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA038217R

(59) Carlo, T.; Sierra, R.; Berget, S. M. *Mol. Cell. Biol.* **2000**, *20*, 3988−3995.
(60) McCullough, A. J.; Berget, S. M. *Mol. Cell. Biol.* **2000**, *20*, 9225−9235.
(61) Friedman, K. A.; Heller, A. *J. Phys. Chem. B* **2001**, *105*, 11859−11865.
(62) Hall, D. B.; Barton, J. K. *J. Am. Chem. Soc.* **1997**, *119*, 5045−5046.
(63) Bernardi, G. *Gene* **2000**, *259*, 31−43.
(64) Eyre-Walker, A.; Hurst, L. D. *Nat. Rev. Genet.* **2001**, *2*, 549−555.
(65) Sueoka, N. *J. Mol. Evol.* **1992**, *34*, 95−114.
(66) Hamilton, M. L.; Guo, Z.; Fuller, C. D.; et al. *Nucleic Acids Res.* **2001**, *29*, 2117−2126.