Conservation of Stem-Loop Potential in Introns of Snake Venom Phospholipase A_2 Genes: An Application of FORS-D Analysis

D. R. Forsdyke
Department of Biochemistry, Queen's University

Venomous snakes and their prey are engaged in an "arms race" in which resistance to venom proteins must be countered by the adaptation of the proteins to overcome that resistance. The toxicity of venoms is contributed to by phospholipase A_2, the genes of which are under extreme positive selection pressure. In contrast to most other gene families, intraspecies and interspecies comparisons of venom phospholipase A_2 genes reveal *low* conservation of exons and *high* conservation of introns. FORS-D analysis (folding of randomized sequence difference analysis) implies that the sequences of introns of venom phospholipase A_2 genes have been under pressure to accept mutations which conserve the potential to form single-strand stem loops, whereas exon sequences have evaded this pressure. These results are consistent with the hypothesis that stem-loop-based recombination evolved in the early "RNA world," and the potential to generate stem loops became widely dispersed throughout genomes. The subsequent conflict between stem-loop potential and protein-encoding potential was resolved by use of synonymous codons, acceptance of conservative amino acid changes to widen the range of codon choice, and encoding proteins in dispersed segments interrupted by regions of high stem-loop potential (introns). In genes under strong Darwinian selection, the pressure on genes to diversify in order to prevent unwanted recombination events is largely accommodated by exons. This leaves introns, which normally predominate in this role, free to accommodate to pressure for stem-loop formation.

Introduction

In a previous paper (Forsdyke 1995*a*), I suggested that, if efficient recombination is evolutionarily advantageous, then it is possible that it would have evolved in the early "RNA world," prior to the emergence of protein-encoding genes (Joyce and Orgel 1993). If this were so, and if stem loops were involved in this primitive recombination process (Romanova et al. 1986; Kleckner et al. 1991; Kleckner and Weiner 1993), then protein-encoding genes would have been *imposed* on sequences which had already been adapted for stem-loop formation. There could then have been a conflict, since a sequence might not be capable of locally optimizing both its base order–determined stem-loop potential and its base order–determined protein-encoding potential. The conflict could have been accommodated in three ways: the use of those synonymous codons which best facilitate stem-loop formation; the use of conservative amino acids to widen the range of codon choice; the encoding of proteins in sections (exons), interspersed with regions of high stem-loop potential (introns).

The latter "introns early" theory states that the early introns would have been regions of high base order–determined stem-loop potential, and it predicts that traces of this primitive arrangement should be detectable in some modern genes. To examine this, FORS-D analysis was applied to various gene sequences. A tendency for base order–determined stem-loop potential to localize to introns was apparent, but in several cases the potential was equally apparent in exons and introns (Forsdyke 1995*a*). What was needed was a gene that had been under such extreme selection pressure for function that accommodation of the potential in exons would have been unlikely.

The snake venom phospholipase A_2 (PLA_2) genes appear ideal for this purpose. Snakes are engaged in an intense "arms race" with their prey (or predators), which can acquire venom resistance (Harding and Welch 1980). Recently, the genomic sequences of PLA_2 genes have become available from habu snakes (Nakashima et al. 1993) and rattlesnakes (John et al. 1994). In keeping with an intense positive selection pressure for change, substitution rates obtained by comparing homologous

Key words: recombination, introns early theory, intron conservation, exon variation, snake venom, phospholipase A_2, FORS-D analysis.

Address for correspondence and reprints: Donald R. Forsdyke, Department of Biochemistry, Queen's University, Kingston, Ontario K7L3N6, Canada; E-mail: forsdyke@qucdn.queensu.ca.

Mol. Biol. Evol. 12(6):1157–1165. 1995.
© 1995 by The University of Chicago. All rights reserved.
0737-4038/95/1206-0019$02.00

exons are high, whereas introns sequences are remarkably conserved. In exons, nonsynonymous substitutions exceed synonymous substitutions.

The present paper describes the application of FORS-D analysis to these sequences. Contrary to the assertion that for venom PLA_2 genes "the regions corresponding to introns of precursor RNAs involved no significant secondary structure" (Nakashima et al. 1993, p. 5966), FORS-D analysis indicates an evolutionary pressure to conserve intron sequences in order to maintain the potential for secondary structure, and hence the potential to undergo recombination.

Methods

FORS-D Analysis

This evaluates base order–determined stem-loop potential, which closely corresponds to the "statistically significant" stem-loop potential of Le and Maizel (1989). The theory and application of the method were described previously (Forsdyke 1995*a*).

Determination of Substitution Density

Sequences of different PLA_2 genes were aligned with the program GENALIGN using the "regions" method, a matching weight of 1.0, and a deletion weight (gap penalty) of −0.5 (Martinez 1988). The program was accessed through the Bionet on-line computing service (IntelliGenetics Inc., Mountain View, California). Base substitutions and "indels" (insertions or deletions) were counted in successive 200-nt windows, each of which overlapped the preceding window by 150 nt. Substitutions dominated the alignments, and changes in deletion weight over a range from −0.25 to −1.0 caused few interchanges between substitutions and indels. Thus, the general patterns of substitution and indel densities were independent of deletion weight. No corrections were made for the possibility of multiple substitutions.

Results

High Substitution Density in Exons

Many species have multiple PLA_2 genes, which usually consist of four exons (Davidson and Dennis 1990). In snakes a subset of these genes encodes proteins which contribute to the toxicity of venom. Intraspecies comparisons of base substitutions in snake venom phospholipase genes show high substitution densities in exons and low substitution densities in introns. This applies to the habu snake (Nakashima et al. 1993) and to the rattlesnake (John et al. 1994). Figure 1*A* demonstrates this, both within species and between species. Substitutions are scored for successive 200-nt windows which overlap by 150 nt. Relative to the rattlesnake PLA_2 acidic subunit gene (GenBank designation U01026), the rattlesnake PLA_2 basic subunit gene (GenBank designation U01027) and two habu snake PLA_2 genes (GenBank designations TFLPA2P1 and TFLPA2P2) show many substitutions in all but the first exon, most of which does not code for protein. In the fourth exon, substitution density is greatest in the protein-encoding region. Because some of the 200-nt windows, centered in introns, overlap exons, sharp demarcations between the low substitution density introns and high substitution density exons are not apparent. However, the relatively low intron substitution density is not uniform in the second intron; here the density increases in the 3′ half of the intron. Essentially similar density patterns are obtained when different venom PLA_2 genes are taken as a basis for comparison (e.g., TFLPA2P2 compared with U01026, U01027 and TFLPA2P1; data not shown). Thus, although the particular bases substituted tend to be characteristic of a particular pair of genes, the overall substitution patterns are unlikely to be random.

Figure 1*B* shows similar data for base insertions and deletions. Indels are in both coding and noncoding regions. Major peaks occur at the 3′ end of the second intron in the cases of the U01026/U1027 and U1026/TFLPA2P2 comparisons. In the former case the peak is associated with a 38-nt CG-rich sequence in U01026, which is also present in the habu snake genes but not in U01027. In the latter case the peak is associated with a microsatellite repeat $(CA)_n$ in TFLPA2P2, which is not so evident in the other snake venom PLA_2 genes. The indel density patterns shown in figure 1*B* are essentially similar to those obtained when different venom PLA_2 genes are taken as a basis for comparison (data not shown). Thus, although the actual bases inserted or deleted tend to be characteristic of a particular pair of genes, the overall patterns are unlikely to be random.

Reciprocal Relationship between Substitution Density and FORS-D

The 200-nt windows used to characterize substitution and indel densities (fig. 1) were subjected to the program FOLD (Zuker 1989) to determine the energetically most favorable folded structures. In figures 2*B* (rattlesnake gene U01026) and 3*B* (rattlesnake gene U01027), the minimum free energies associated with each structure (FONS values) are shown together with the mean values for 10 randomized versions of the same sequence windows (FORS-M values). In keeping with previous observations on a variety of genes (Forsdyke 1995*a*), FONS values tend to be more negative than FORS-M values (indicating greater stability of the secondary structure of the natural sequence). This is particularly apparent in the 5′ noncoding part of exon 1, in the first part of intron 2, and in intron 3. The differences (FORS-M less FONS) are expressed in figures 2*A* and

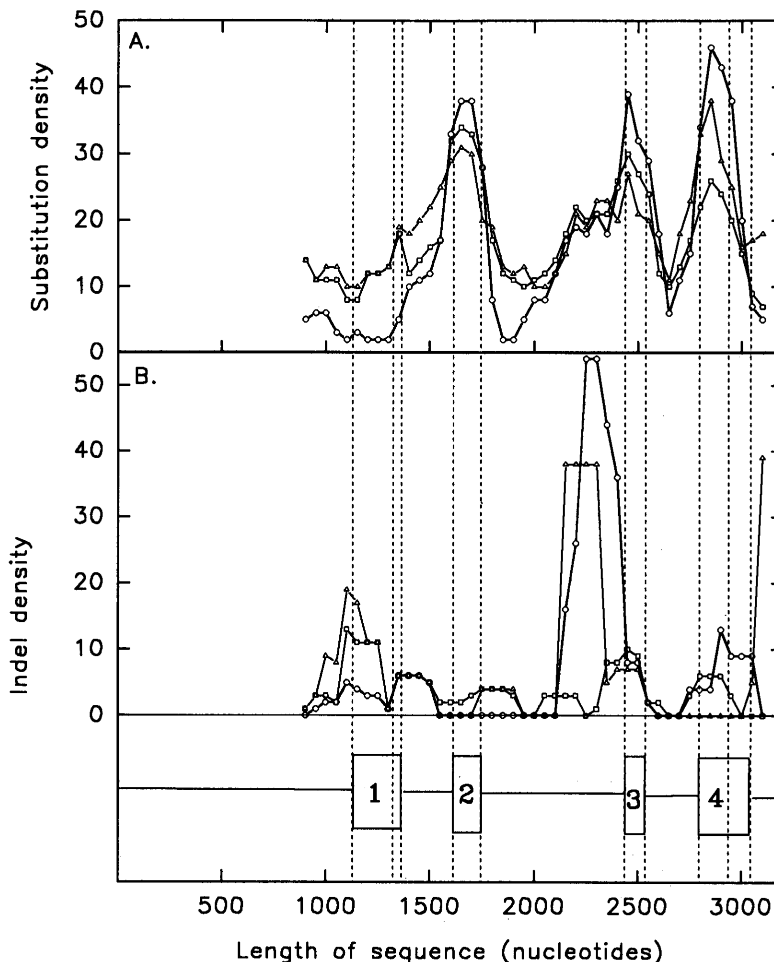

FIG. 1.—Distribution of (*A*) substitutions and (*B*) indels in rattlesnake PLA_2 acidic subunit gene U01026, relative to rattlesnake PLA_2 basic subunit gene U01027 (*circles*), habu snake PLA_2 gene TFLPA2P1 (*squares*), and habu snake PLA_2 gene TFLPA2P2 (*triangles*). Substitutions and indels are counted in successive 200-nt windows, each of which overlaps the preceding window by 150 nt. Data points at 50-nt intervals each correspond to the center of a window. The positions of the four exons are shown as boxes in *B*. Vertical dashed lines indicate exon borders, the beginning of the protein-coding part of exon 1 and the end of the protein-coding part of exon 4.

3*A* as the FORS-D (folding of randomized sequence difference) values. Positive FORS-D values reflect a contribution of base order (primary sequence), rather than of base composition, to the stability of the optimum structure. The previously noted tendency for protein-encoding parts of exons to have zero or negative FORS-D values and for introns to have positive FORS-D values (Fordyke 1995*a*) is generally apparent. In the case of U01026 the first intron has low values, but this is not seen in the case of U01027. In the case of U01027 the coding part of the fourth exon corresponds to a relative decline in FORS-D value but remains positive. It should be noted that high FORS-D values, reflecting local adaptation of base order for the generation of secondary structure, are not necessarily correlated with high FONS values (e.g., high FORS-D values occur in the 3′ half of the second intron and in the 3′ noncoding part of the fourth exon).

Figures 2*C*, 2*D*, 3*C*, and 3*D* show FONS, FORS-M, and FORS-D values for U01026 and U01027 (rattlesnake) cDNAs. Fine vertical lines indicate sites of intron removal. FORS-D values are low in the regions corresponding to the middle two exons but increase in the 5′ and 3′ exons, which contain extensive noncoding regions.

Figures 2*A* and 3*A* also show substitution and indel densities taken from figure 1. There is an approximately reciprocal relationship between FORS-D values and substitution density. Where FORS-D values are low (coding parts of exons), substitution densities are high. Where FORS-D values are high (introns; 5′ and 3′ noncoding regions), substitution densities are low. However, the 3′ end of the second intron, where indel density is highest, is characterized by relatively high values both for FORS-D and substitution density. Similar results were obtained when

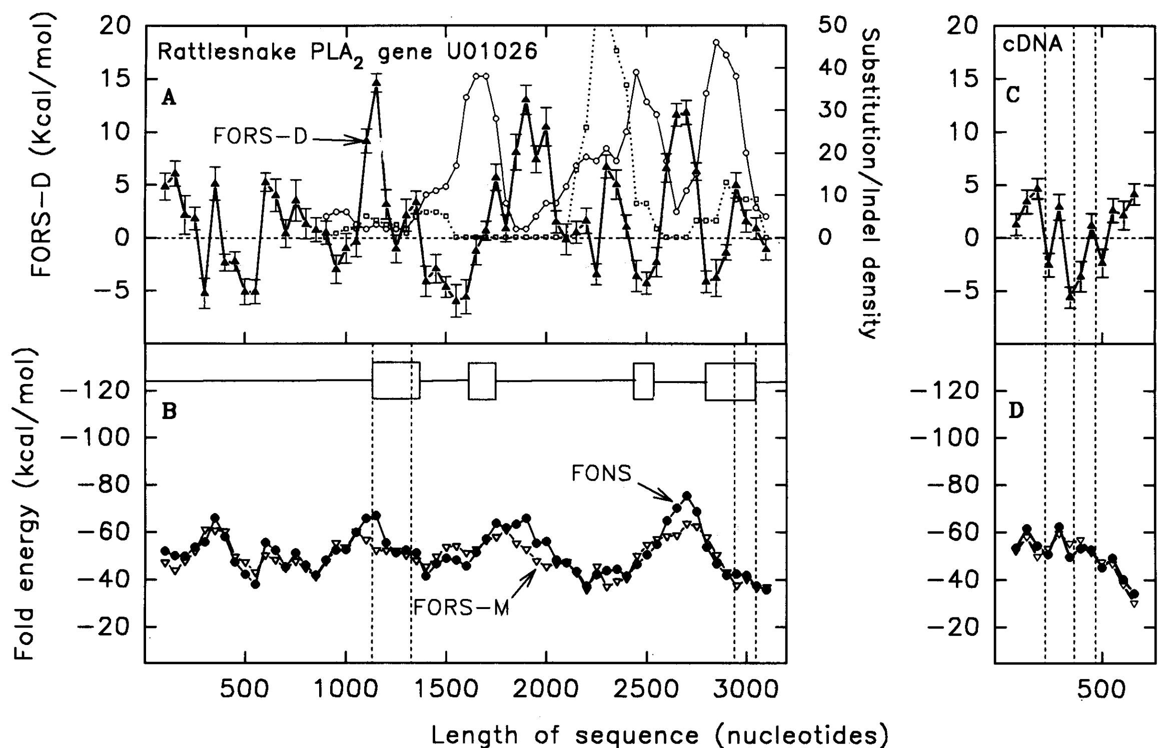

FIG. 2.—Comparison of distribution of values for (B) FONS (*solid circles*), FORS-M (*open triangles*), and (A) FORS-D (*solid triangles*), with distribution of base substitutions (*open circles*) and indels (*open squares*), for the rattlesnake PLA_2 acidic subunit gene U01026. Values are for the same 200-nt windows as shown in fig. 1. Substitutions and indels are relative to the rattlesnake PLA_2 basic subunit gene U01027. Boxes in *B* indicate the location of the four exons, with four fine vertical lines showing the beginning of exon 1, the beginning of the protein-coding part of exon 1, the end of the protein-coding part of exon 4, and the end of exon 4. On the right are shown (D) FONS, FORS-M, and (C) FORS-D values for the cDNA. Vertical lines indicate splice sites (exon-intron junctions).

the other venom PLA_2 genes under study were compared in this manner.

More objective evidence for a relationship between secondary structure and substitution density was provided by plotting individual window values for FONS, FORS-M, and FORS-D against the corresponding substitution densities. For each pair of genes three linear regression plots were obtained. Figure 4 shows results for three gene pairs (U01026/U01027, U01027/U01026, U01026/TFLPA2P2). In general, plots for FORS-M are horizontal, plots for FONS slope down slightly, and plots for FORS-D slope down more steeply. Within each figure are listed values for the slope (s), correlation coefficient (r), and the probability (P) that the slope is not significantly different from zero. The latter two values are most significant (e.g., high values for r and low values for P) in the case of plots of FORS-D against substitution density and least significant in the case of plots of FORS-M against substitution density. The results of similar plots for other gene combinations are shown in table 1. Although values for some gene combinations have relatively low r values and high P values (e.g., fig. 4*G*, 4*H*, 4*I*), direct inspection of the corresponding plots of values as a function of sequence position (e.g. fig. 3*A*) tends to support the generalization that, in most parts of a gene, there is a reciprocal relationship between substitution density and FORS-D value.

Discussion

Although unusual, high conservation of noncoding sequences in cDNAs, to an extent exceeding that of neighboring coding sequences, has been observed both within species (Blum et al. 1990) and between species (Duret et al. 1993). Ogawa and co-workers (1992) compared PLA_2 cDNAs of the habu snake (*Trimeresurus flavoviridis*), which were found to have extremely high conservation of 5′ and 3′ noncoding sequences (98% and 89% identities, respectively). They correlated this with high stem-loop potential and suggested an important functional role at the translation level. They also noted high conservation of the nucleotide sequence corresponding to the 16 amino acid signal peptide. However, the remaining protein-encoding sequence showed low conservation (67% identity), mainly due to nucleotide substitutions. An unusually high proportion of these were nonsynonymous, thus changing the protein se-

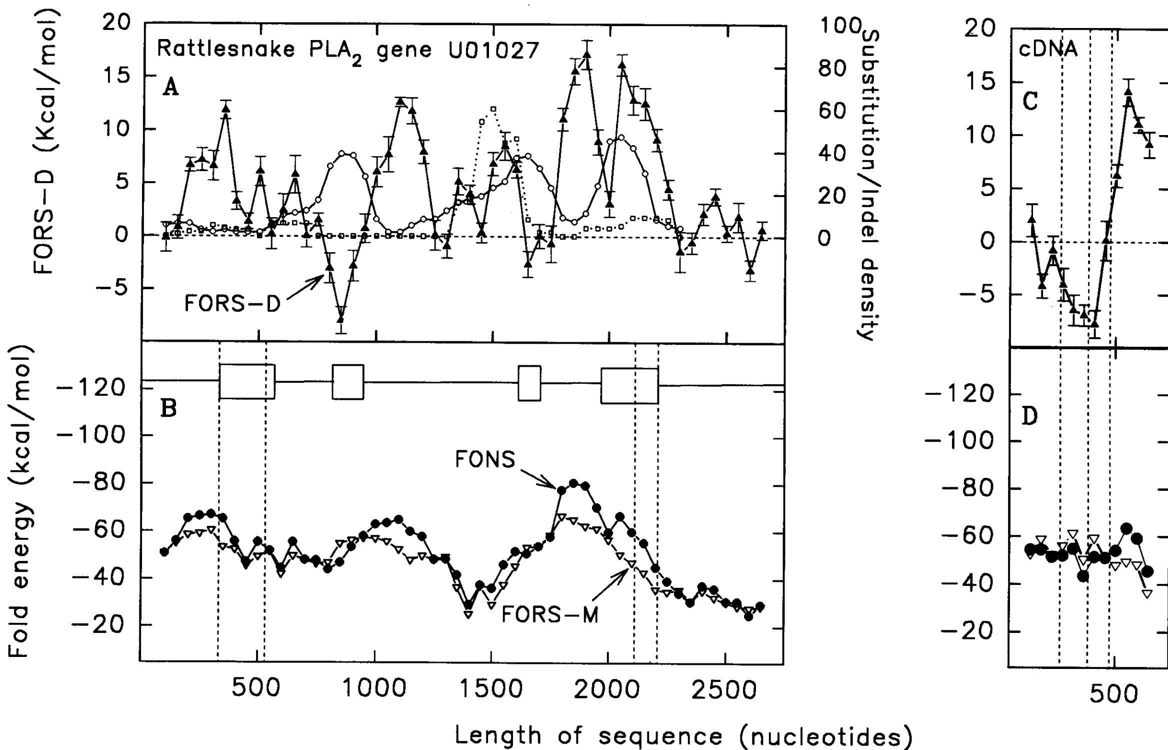

FIG. 3.—Comparison of distribution of values for FONS, FORS-M, and FORS-D, with distribution of base substitutions and indels, for the rattlesnake PLA_2 basic subunit gene U01027. Substitutions and indels are relative to the rattlesnake PLA_2 acidic subunit gene U01026. Symbols and other details are as in fig. 2.

quence. These observations were consistent with strong selection pressures both to *conserve* the signal peptide function and to *change* the function of the mature protein (122 amino acids). The change would be required to counter adaptations to resistance of the prey. The signal peptide would, of course, be discarded and would not interact with the prey. PLA_2 generates a cytotoxic product (Takeda et al. 1982) and reacts with broad specificity binding proteins (Lambeau et al. 1994), which might adapt to generate resistance.

A similar evolutionary process (positive selection) appears to operate in the cases of (1) serine proteinase inhibitor genes, the products of which may need to counter adaptations to resistance of parasite proteases (Hill and Hastie 1987); (2) major histocompatibility genes, the products of which may need to counter changes in peptide sequences in the proteins of parasites (Hughes and Nei 1989; Forsdyke 1991, 1994); and (3) retroviral genes, which may need to counter host immune defenses (Scpaer and Mullins 1993).

When habu snake PLA_2 genomic sequences became available, Nakashima and co-workers (1993) found that intron sequences were as highly conserved as the 5′ and 3′ noncoding regions. This made explanations of functional effects at the translation level unlikely and raised the possibility of effects of nucleic acid secondary structure at the genomic level, or at the level of RNA processing. However, Nakashima and co-workers found no secondary structure when the sequences were analyzed by the method which had been used to show secondary structure in 5′ and 3′ noncoding regions (Zuker and Steigler 1981; Ogawa et al. 1992).

The observations on habu snake venom PLA_2 genes were also found to apply to the venom PLA_2 genes of the Mojave rattlesnake (*Crotalus scutulatus scutulatus*). John and co-workers (1994) pointed out that "the greater intron identity appears to violate evolutionary dogma, which predicts greater divergence in the non-coding regions" (p. 231). They suggested a critical role of introns in pre-mRNA stabilization or processing.

The present work began by demonstrating similar patterns of conservation, whether comparisons were made within species or between species (fig. 1). The patterns were similar for different gene pairs, but the actual changes tended to be characteristic of a particular gene pair. Thus, although gene duplication generating multiple-venom PLA_2 genes may have occurred in an ancestor prior to the divergence of the habu snake and rattlesnake lineages (Davidson and Dennis 1990), most sequence differentiation is likely to have occurred within species, after the divergence.

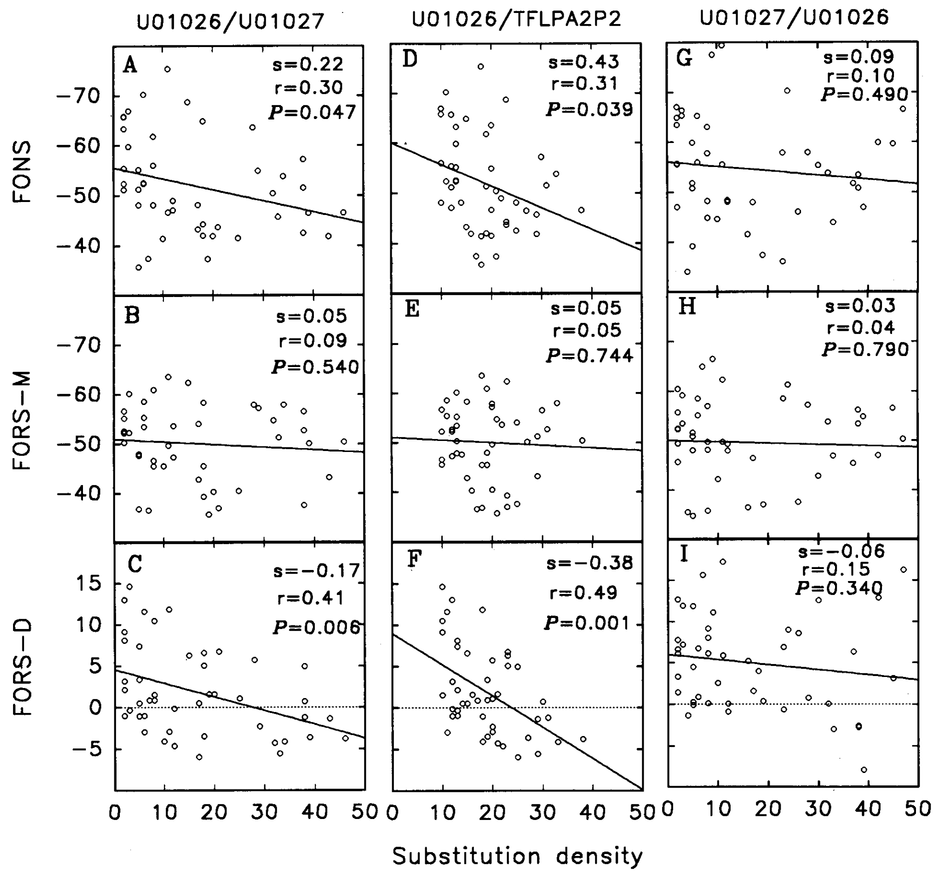

FIG. 4.—Linear regression analysis of the relationship between FONS, FORS-M, and FORS-D values (kcal/mol) and substitution density (base exchanges/200-nt window) in snake venom PLA_2 genes. For the left column of figures (*A, B, C*), data are from fig. 2 (FONS, FORS-M, and FORS-D values for rattlesnake PLA_2 acidic subunit gene U01026, and substitutions relative to rattlesnake PLA_2 basic subunit gene U01027). The middle column (D,E,F) shows FONS, FORS-M and FORS-D values for rattlesnake PLA_2 acid subunit gene U01026 and substitutions relative to habu snake PLA_2 gene TFLPA2P2. For the right column (*G, H, I*), data are from fig. 3 (FONS, FORS-M, and FORS-D values for rattlesnake PLA_2 basic subunit gene U01027, and substitutions relative to rattlesnake PLA_2 acidic subunit gene U01026). Parameters of the least-squares line shown in each figure are slope (s), the correlation coefficient (r), and the probability that the slope of the line is not significantly different from zero (P).

Similarity of patterns of different species indicates similar selective constraints.

Direct observations of the folding into stem-loop structures of sections of the sequences (FONS values) indicated some increased secondary structure potential in introns, relative to exons (figs. 2*B*, 3*B*). This was particularly apparent in the third intron. The failure of others to note this may reflect the use in the present work of an improved folding algorithm (Zuker 1989) and better energy values for base stacking and loop destabilization (Turner et al. 1988). However, FONS values reflect the contribution of both base order and base composition. Base composition probably serves genome-wide or genome-sector specific functions rather than gene-specific functions, as discussed elsewhere (Forsdyke 1995*a*, 1995*b*). Evolutionary effects on the base order became apparent when the negative FONS values (greater negativity reflecting more stable structures) were subtracted from the negative FORS-M values (which reflect the contribution of base composition alone). This produced FORS-D values. Positive FORS-D values, indicating the acceptance of mutations which contribute to the stability of secondary structure by changing base order, were generally present both in introns and in 5′ and 3′ noncoding regions. Thus, the structures in these regions are likely to be functionally relevant, although they may not necessarily be correlated with high FONS values. Zero and negative FORS-D values, reflecting no acceptance of such mutations, were generally present in the exons encoding the mature PLA_2 protein (figs. 2*A*, 3*A*). Thus, the conservation of intron sequences could reflect a functional constraint for which base order–determined stem-loop potential is required. The direct relationship between base order–determined stem-loop

Table 1
Summary for Venom Phospholipase A_2 Genes of Linear Regression Analyses of FONS, FORS-M, and FORS-D Values as a Function of Substitution Density

		Rattlesnake						Habu Snake					
		U01026			U01027			TFLPA2P1			TFLPA2P2		
Snake GenBank Name		*s*	*r*	*P*	*s*	*r*	*P*	*s*	*r*	*P*	*s*	*r*	*P*
Rattlesnake:													
U01026	FONS				0.22	0.30	0.047	0.32	0.24	0.112	0.43	0.31	0.039
	FORS-M				0.05	0.09	0.540	0.01	0.01	0.946	0.05	0.05	0.744
	FORS-D				−0.17	0.41	0.006	−0.31	0.42	0.004	−0.38	0.49	0.001
U01027	FONS	0.09	0.10	0.490				0.13	0.17	0.244	0.02	0.02	0.920
	FORS-M	0.03	0.04	0.790				0.11	0.18	0.222	−0.09	0.10	0.504
	FORS-D	−0.06	0.15	0.340				−0.02	0.06	0.692	−0.10	0.19	0.216
Habu snake:													
TFLPA2P1	FONS	0.28	0.22	0.140	0.02	0.06	0.693				0.26	0.23	0.132
	FORS-M	−0.03	0.03	0.817	0.01	0.02	0.908				0.04	0.05	0.757
	FORS-D	−0.32	0.47	0.001	−0.02	0.09	0.550				−0.22	0.35	0.017
TFLPA2P2	FONS	0.10	0.06	0.703	0.01	0.01	0.940	0.21	0.15	0.336			
	FORS-M	0.02	0.01	0.932	−0.02	0.02	0.889	0.08	0.08	0.618			
	FORS-D	−0.08	0.11	0.478	−0.03	0.06	0.705	−0.13	0.20	0.182			

NOTE.—Linear regression analyses, examples of which are shown in fig. 4, were performed for various pairs of snake venom PLA_2 genes. FONS, FORS-M, and FORS-D values were obtained for the vertically listed genes. Substitution densities were relative to the horizontally listed genes. For a particular gene pair and either the FONS, FORS-M, or FORS-D value, the set of three numbers (*s*/*r*/*P*) refers to the slope of the linear regression plot (*s*), the correlation coefficient (*r*), and the probability that the slope is not significantly different from zero (*P*).

potential and sequence conservation (i.e., inverse relationship to substitution density) was shown by linear regression analysis (fig. 4; table 1). Similar observations have been made on retroviral (HIV-1) sequences (Le et al. 1988, 1989; and unpublished data).

The functional constraint for which base order–determined stem-loop potential is required could reflect some role at the pre-mRNA level. However, as set out previously (Forsdyke 1995*a*), a role of stem loops in recombination is the preferred explanation. In most genes some compromise between the encoding of base order–determined stem-loop potential and protein potential is achieved; base order–determined stem-loop potential, while often greater in introns, is also found in exons, particularly large exons. Indeed, exons may be large *because* they have been able to achieve a satisfactory compromise. Otherwise they would be small. The intense *externally* driven selection pressure for change acting on exons encoding the mature PLA_2 protein would have been accompanied by an equally intense *internally* driven selection pressure for conservation of base order–determined stem-loop potential, acting on introns. Genes which mutate to decrease base order–determined stem-loop potential would be less able to undergo recombination and hence would be less able to repair DNA damage (Bernstein and Bernstein 1991).

If conservation of base order–determined stem-loop potential is so important, why do introns in general not show better conservation (Forsdyke 1995*a*)? One explanation is that in some genes exon sequences can compromise and retain some base order–determined stem-loop potential, thus placing less demand on introns in this respect. However, just as exons may be under intense selection pressure either to remain constant (e.g., venom PLA_2 signal peptide-encoding exons) or to change (e.g., mature venom PLA_2-encoding exons), so introns may be subject to similar divergent pressures. Introns "want" to conserve their unique pattern of stem-loop potential in order to facilitate *intra*species recombination, but they also "want" to change their unique pattern in order to avoid *inter*species recombination (a basis of speciation, as discussed elsewhere; Forsdyke 1995*a*, 1995*b*). This would be particularly important in the case of genes with highly conserved exons.

Indels probably play an important role in interfering with the initial alignment (register) of stem loops in potentially recombining homologous strands (Tomizawa 1993). Although within a genome the introns for different venom PLA_2 genes may show great conservation (tending to provoke recombination), indels would make it more likely that recombination would occur only between identical genes. The differences in exon sequences would also ensure that, even if the initial "kissing" interaction between the loops of stem loops were stable (Kleckner and Weiner 1993; Tomizawa 1993), subsequent consummation of the

union would be unlikely (Rayssiguier et al. 1989; Radman and Wagner 1993).

The results presented here are consistent with the new "introns-early" hypothesis (Forsdyke 1995*a*). It is becoming increasingly evident that introns interrupt nucleic acid information *in general,* not just protein-encoding information. Thus, introns are often found in 5′ and 3′ noncoding regions (Hawkins 1988). Genes encoding "mRNAs" which have no protein product have been identified; the gene sequences are interrupted by introns just like normal protein-encoding genes (Brannan et al. 1990; Brockdorff et al. 1992). The latter observation, and those presented in this paper, reinforce the viewpoint that there is generally no significant association between intron locations and features of protein structure (Stoltzfus et al. 1994).

The results presented here also suggest that it would be of considerable interest to extend the analysis to other genes under positive selection pressure (Hill and Hastie 1987; Hughes and Nei 1989), as has been done in the case of retroviral genomes (unpublished data). Indeed, recent studies show that regions encoding peptide-binding sites in MHC genes have negative FORS-D values (unpublished data). It has been noted that troponin-C genes are under positive Darwinian selection (Ohta 1994) and that certain exons show very marked negative FORS-D values (Forsdyke 1995*a*). It would also be of interest to compare venom PLA_2 genes with other PLA_2 genes which would not be under such pressure.

Acknowledgments

This work was supported by a grant from the Medical Research Council of Canada.

LITERATURE CITED

BERNSTEIN, C., and H. BERNSTEIN. 1991. Aging, sex and DNA repair. Academic Press, San Diego.

BLUM, S., R. E. FORSDYKE, and D. R. FORSDYKE. 1990. Three human homologs of a murine gene encoding an inhibitor of stem cell proliferation. DNA Cell Biol. **9**:589–602.

BRANNAN, C. I., E. C. DEES, D. S. INGRAM, and S. M. TILGHMAN. 1990. The product of the H19 gene may function as an RNA. Mol. Cell. Biol. **10**:28–36.

BROCKDORFF, N., A. ASHWORTH, G. F. KAY, V. M. MCCABE, D. P. NORRIS, P. J. COOPER, S. SWIFT, and S. RASTAN. 1992. The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell **71**:515–526.

DAVIDSON, F. F., and E. A. DENNIS. 1990. Evolutionary relationships and implications for the regulation of phospholipase A_2 from snake venom to human secreted forms. J. Mol. Evol. **31**:228–238.

DURET, L., F. DORKELD, and C. GAUTIER. 1993. Strong conservation of non-coding sequences during vertebrate evolution: potential involvement in post-transcriptional regulation of gene expression. Nucleic Acids Res. **21**:2315–2322.

FORSDYKE, D. R. 1991. Early evolution of MHC polymorphism. J. Theor. Biol. **150**:451–456.

———. 1994. Relationship of X chromosome dosage compensation to intracellular self/not-self discrimination: a resolution of Muller's paradox? J. Theor. Biol. **167**:7–12.

———. 1995*a*. A stem-loop "kissing" model for the initiation of recombination and the origin of introns. Mol. Biol. Evol. **12**:949–958.

———. 1995*b*. Relationship between the hierarchy of frequencies of complementary trinucleotide pairs and percentage G+C. J. Mol. Evol. (in press).

HARDING, K. A., and K. R. G. WELCH. 1980. Venomous snakes of the world. Pergamon, Oxford.

HAWKINS, J. D. 1988. A survey of intron and exon lengths. Nucleic Acids Res. **16**:9893–9905.

HILL, R. E., and N. D. HASTIE. 1987. Accelerated evolution in the reactive center regions of serine protease inhibitors. Nature **326**:96–99.

HUGHES, A. L., and M. NEI. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA **86**: 958–962.

JOHN, T. R., L. A. SMITH, and I. I. KAISER. 1994. Genomic sequences encoding the acidic and basic subunits of Mojave toxin: unusually high sequence identity of non-coding regions. Gene **139**:229–234.

JOYCE, G. F., and L. E. ORGEL. 1993. Prospects for understanding the origin of the RNA world. Pp. 1–25 *in* R. F. GESTELAND and J. F. ATKINS, eds. The RNA world. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

KLECKNER, N., R. PADMORE, and D. K. BISHOP. 1991. Meiotic chromosome metabolism: one view. Cold Spring Harbor Symp. Quant. Biol. **56**:729–743.

KLECKNER, N., and B. M. WEINER. 1993. Potential advantages of unstable interactions for pairing of chromosomes in meiotic, somatic and premeiotic cells. Cold Spring Harbor Symp. Quant. Biol. **58**:553–565.

LAMBEAU, G., P. ANCIAN, J. BARHANIN, and M. LAZDUNSKI. 1994. Cloning and expression of a membrane receptor for secretory phospholipase A_2. J. Biol. Chem. **269**:1575–1578.

LE, S.-Y., J.-H. CHEN, M. BRAUN, M. GONDA, and J. V. MAIZEL. 1988. Stability of RNA stem-loop structure and distribution of non-random structure in the human immunodeficiency virus (HIV-1). Nucleic Acids Res. **16**:5153–5168.

LE, S.-Y., J.-H. CHEN, D. CHATTERJEE, and J. V. MAIZEL. 1989. Sequence divergence and open regions of RNA secondary structures in the envelope regions of the 17 human immunodeficiency virus isolates. Nucleic Acids Res. **17**: 3275–3288.

LE, S.-Y., and J. V. MAIZEL. 1989. A method for assessing the statistical significance of RNA folding. J. Theor. Biol. **138**: 495–510.

MARTINEZ, H. M. 1988. A flexible multiple sequence alignment program. Nucleic Acids Res. **16**:1683–1691.

NAKASHIMA, K.-I., T. OGAWA, N. ODA, M. HATTORI, Y. SAKAKI, H. KIHARA, and M. OHNO. 1993. Accelerated evolution of *Trimeresurus flavoviridis* venom gland phospholipase A_2 isozymes. Proc. Natl. Acad. Sci. USA **90**:5964–5968.

OGAWA, T., N. ODA, K.-I. NAKASHIMA, H. SASAKI, M. HATTORI, Y. SAKAKI, H. KIHARA, and M. OHNO. 1992. Unusually high conservation of untranslated sequences in cDNAs for *Trimeresurus flavoviridis* phospholipase A_2 isoenzymes. Proc. Natl. Acad. Sci. USA **89**:8557–8561.

OHTA, T. 1994. Further examples of evolution by gene duplication revealed through DNA sequence comparisons. Genetics **138**:1331–1337.

RADMAN, M., and R. WAGNER. 1993. Mismatch recognition in chromosomal interactions and speciation. Chromosoma **102**:369–373.

RAYSSIGUIER, C., D. THALER, and M. RADMAN. 1989. The barrier to recombination between *Echerichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. Nature **342**:396–401.

ROMANOVA, L. I., V. M. BLINOV, E. A. TOLSKAYA, E. G. VIKTOROVA, M. S. KOLESNIKOVA, E. A. GUSEVA, and V. I. AGOL. 1986. The primary structure of crossover regions of intertypic poliovirus recombinants: a model of recombination between RNA genomes. Virology **155**:202–213.

SCPAER, E. G., and J. I. MULLINS. 1993. Rates of amino acid change in the envelope protein correlates with pathogenicity of primate lentiviruses. J. Mol. Evol. **37**:57–65.

STOLTZFUS, A., D. F. SPENCER, M. ZUKER, J. M. LOGSDON, and W. F. DOOLITTLE. 1994. Testing the exon theory of genes: the evidence from protein structure. Science **265**: 202–207.

TAKEDA, A., R. G. E. PALFREE, and D. R. FORSDYKE. 1982. Role of serum in inhibition of cultured lymphocytes by lysophosphatidylcholine. Biochem. Biophys. Acta **710**:87–98.

TOMIZAWA, J. 1993. Evolution of functional structures of RNA. Pp. 419–445 *in* R. F. GESTELAND and J. F. ATKINS, eds. The RNA world. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

TURNER, D. H., N. SUGIMOTO, and S. M. FREIER. 1988. RNA structure prediction. Annu. Rev. Biophys. Chem. **17**:167–192.

ZUKER, M. 1989. Computer prediction of RNA secondary structure. Meth. Enzym. **180**:262–289.

ZUKER, M., and P. STEIGLER. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133–148.

JAN KLEIN, reviewing editor

Received February 16, 1995

Accepted May 19, 1995