ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Feb. 1994, p. 268–274
0066-4804/94/$04.00+0

Vol. 38, No. 2

Antiviral Drug Resistance Mutations in Human Immunodeficiency Virus Type 1 Reverse Transcriptase Occur in Specific RNA Structural Regions

RAYMOND F. SCHINAZI,[1,2]* ROBERT M. LLOYD, JR.,[1,2] CHANDRA S. RAMANATHAN,[3] AND ETHAN WILL TAYLOR[3]

Veterans Affairs Medical Center (Atlanta), Decatur, Georgia 30033[1]; Laboratory of Biochemical Pharmacology, Department of Pediatrics, Emory University School of Medicine, Atlanta, Georgia 30322[2]; and Computational Center for Molecular Structure and Design and Department of Medicinal Chemistry, The University of Georgia, Athens, Georgia 30602[3]

Received 30 August 1993/Returned for modification 27 October 1993/Accepted 24 November 1993

A statistically significant correlation exists between the locations of drug resistance mutations (DRMs) observed for various reverse transcriptase inhibitors and features of the secondary structure predicted for the RNA coding for human immunodeficiency virus type 1 reverse transcriptase. The known DRMs map onto "unstable" bases, which are predominantly nonhelical regions (i.e., loops, bulges, and bends) of the predicted RNA secondary structure, whereas codons for the key conserved residues of polymerase sequence motifs map onto "stable" paired bases involved in helical regions. On the basis of these results, we hypothesize that the secondary structure of the RNA template (in this case, the reverse transcriptase gene itself) may be a previously unrecognized factor contributing to base misincorporation errors during reverse transcription and that, rather than being randomly distributed, mutations are more likely to occur in specific regions of the genome. The results suggest that these "mutation-prone" regions can be predicted by using a standard algorithm for RNA secondary structure.

The emergence of drug-resistant strains of human immunodeficiency virus type 1 (HIV-1) is a problem observed with almost all the current selective inhibitors of HIV-1 reverse transcriptase (RT) (35, 38). It has been widely accepted that the high frequency of base incorporation errors observed for RNA-dependent polymerases like RT is primarily due to the lack of a "proofreading" 3′,5′-exonuclease activity. However, various other factors may also contribute to the high mutation rate that underlies the rapid evolution of exogenous retroviruses, which facilitates the development of drug resistance. Drug resistance mutations (DRMs) produced by these mechanisms are selected because they produce modifications at the protein structure level that preserve function in the presence of drug (by removing a favorable interaction or introducing a steric block to the inhibitor), giving rise to viral "escape" mutants. In the presence of RT inhibitors, viruses with single or multiple mutations that confer resistance to the inhibitor(s) have a replicative advantage and become the predominant population.

Although errors produced during reverse transcription are known to accumulate throughout the genome on an evolutionary time scale (which is accelerated for retroviruses), on a more microscopic level, the local distribution of these mutations has generally been assumed to be random and has rarely been closely examined. We investigated the possibility that there may be inherently "mutation-prone" regions of the retroviral genome associated with structural features at the RNA level by mapping all documented mutations associated with resistance to HIV-1 RT inhibitors (referenced in Table 1) onto the computed RNA secondary structure of the HIV-1 RT-coding region (see Fig. 1 and 2). Statistical analysis of the predicted RNA structures suggested that, overall, they are significantly more thermodynamically stable than what would be expected for random nucleotide sequences of the same size and base composition. Although such structural predictions have well-understood limitations (discussed below), the relationship between the location of DRMs and the predicted RNA structure is striking (Fig. 1 and 2).

Our results suggest that DRMs in the coding region of the HIV-1 RT gene are consistently associated with what we have chosen to call "plateau" regions of the predicted RNA secondary structure (see Fig. 1), which correspond to loops, bulges, or bends in the more conventional representation of the RNA secondary structure (see Fig. 2). For the purposes of the following statistical analysis, it is useful to classify these apparently mutation-prone regions into three classes: (i) unpaired bases, (ii) paired bases that border on an unpaired or unstacked region ("border" bases), and (iii) paired bases involved in weak G-U base pairs. In contrast, regions coding for amino acids essential for RT activity, i.e., the conserved residues of polymerase sequence motifs (30), are predominantly associated with "stable" paired bases (bases in A-U or G-C base pairs that are stacked between paired bases on both sides) that form the helical stems in the secondary structure (see Fig. 2).

MATERIALS AND METHODS

Prediction and statistical analysis of RNA secondary structure. The RNA sequence of interest was the HIV-1 RT-coding region, corresponding to bases 2132 to 3103 of the HIV-1 BRU sequence (42), GenBank accession number K02013. RNA folding (secondary structure) was predicted by using the FOLD program (44, 45) with updated free energy parameters, as implemented in the GCG software package (8).

The potential significance of such predicted structures can

* Corresponding author. Mailing address: Veterans Affairs Medical Center, Medical Research-151, 1670 Clairmont Road, Decatur, Georgia 30033. Phone: (404) 728-7711. Fax: (404) 728-7726.

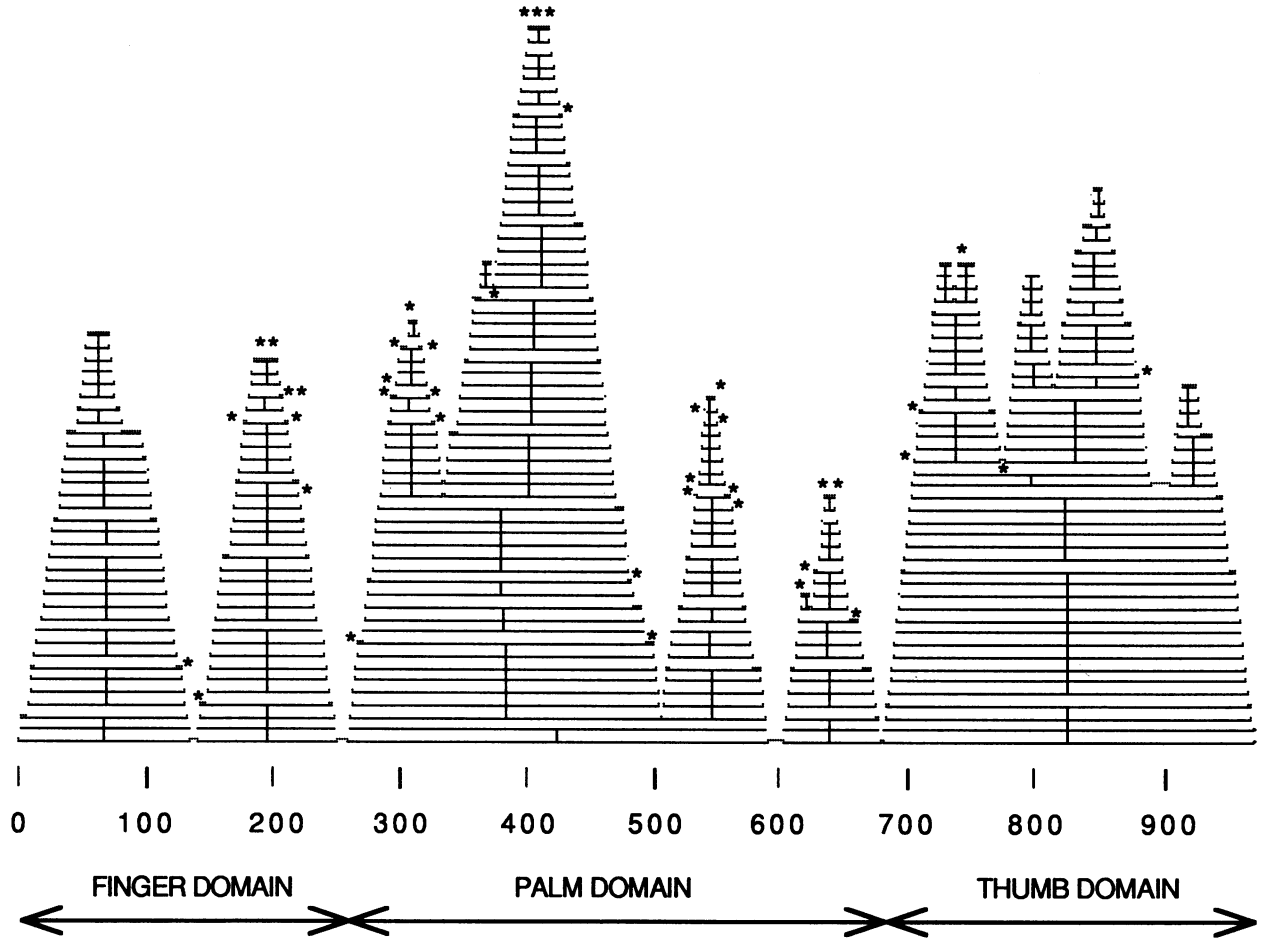

FIG. 1. MOUNTAINS rendition (see text) of the predicted minimum free energy RNA secondary structure for the HIV-1 RT-coding region, with base numbers on the abscissa. The locations of the documented DRMs listed in Table 1 are indicated by asterisks. These mutation sites are predominantly located on plateau regions, most commonly at or near the ends of stem-loop structures (tops of mountains); when not near the top, they are nonetheless consistently on or near bulges in the structure (side plateaus on the mountains). In this rendition of RNA secondary structure, bases are represented by single dots; horizontal lines connect base pairs that form stems; plateaus of horizontal dots represent loops or bulges formed by unpaired bases. Vertical lines represent the stems (helical regions) formed by paired bases; branching is shown by two vertical lines coming up from a horizontal line with a single vertical line below it. Height on the *y* axis is proportional to the number of paired bases between a given location and the base of a stem-loop structure.

be assessed by measuring the deviation of the free energy of stabilization of the predicted structure from the average calculated for randomized structures of identical size and base composition (22). In the present study, the average computed free energy and its standard deviation (SD) were determined from foldings of 100 randomized versions of each sequence of interest. The difference between the computed free energy of folding and the mean value calculated for the randomized sequences, expressed as the number of SD below the average for the randomized sequences, is a useful metric for assessing the relative stability of the computed folding of the actual sequence relative to that attainable by similarly composed random sequences. Since the computed energy values for randomized sequences are approximately normally distributed, an assumption that has been confirmed by statistical analysis (22), a probability (*P* value) can be assigned to the energy value for the actual sequence on the basis of the area under one side of the normal curve beyond the determined SD (*z*) value. This *P* value corresponds to the probability that a random sequence of identical size and base composition would yield an equally stable predicted secondary structure.

RNA secondary structure diagrams. To illustrate several different aspects of the relationship between the computed secondary structure and the locations of DRMs, two different renditions of RNA structure have been used. Figure 1 shows a "MOUNTAINS" rendition (8, 15) of the predicted minimum free energy RNA secondary structure for the HIV-1 RT-coding regions corresponding to the finger, palm, and thumb domains of the RT crystal structure (17), shown as base numbers 1 to 972 on the abscissa, equivalent to the HIV-1 sequence from positions 2132 to 3103 (42). We classified the entire central RT region as "palm," although a small piece within this domain interacts with the N-terminal domain, forming the "fingers" in the crystal structure (17). Individual "mountains" correspond to relatively stable internally H-bonded structures; these structures may also engage in higher-order tertiary interactions with each other. Figure 2 shows a more conventional "SQUIGGLES" rendition (8, 29) of the predicted folding of the RNA for the coding regions of the three major functional domains of HIV-1 RT. These structures correspond precisely to the structures shown in the MOUN-

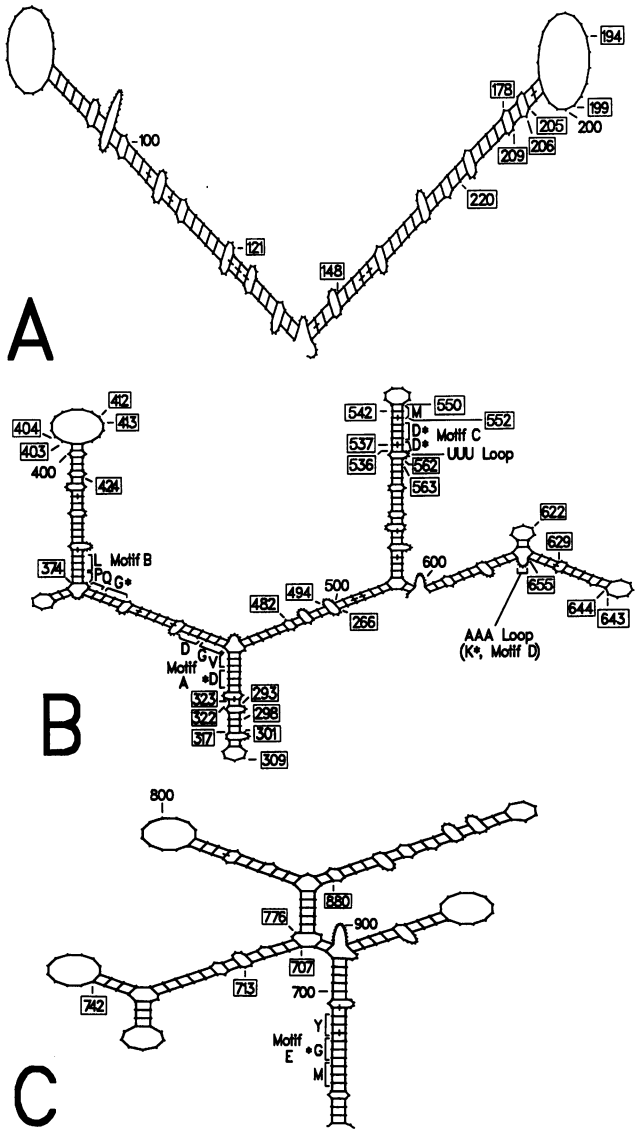

FIG. 2. SQUIGGLES rendition (see text) of the predicted folding of the RNA for the coding regions of the three major functional domains of HIV-1 RT: fingers (A; bases 1 to 255), palm (B; bases 256 to 680), and thumb (C; bases 681 to 972). Bases are numbered every 100 residues (unboxed numbers); boxed numbers correspond to the DRMs that are shown as asterisks in Fig. 1. Locations of the presumably less favorable G-U base pairs are indicated by a slash through the line representing the base pair (e.g., for mutation 552). With the exception of motif D, regions coding for key amino acids of the five RNA-dependent polymerase sequence motifs (30) map onto extended helical regions. Codons for the six invariant amino acids associated with polymerase motifs A to E are indicated by asterisks. The ladder-like regions represent helical stems, with each step being a base pair; unpaired bases are shown as tics on the loop regions. Boxed numbers represent bases involved in DRMs, which are listed and referenced in Table 1.

TAINS rendition of Fig. 1, but provide greater detail and more precise base numbering.

Statistical assessment of the distribution of DRMs relative to RNA structural features. The significance of the observed distribution can be precisely assessed by using the binomial distribution, a discrete probability distribution commonly used for computing the probability of various samplings from a population that can be divided into two distinct sets or outcomes (e.g., paired or unpaired bases). The probability that a randomly sampled individual (a nucleotide base in this case) would fall into one of the two categories would clearly correspond to the percentage of that type (e.g., paired bases) in the entire population. For example, 57% of the bases in the entire RT-coding region are predicted to be paired, so the population fraction $f = 0.57$ is the probability that a randomly picked base will be paired in the predicted structure, and P (unpaired) $= 1 - f$. In a sample of size n, the probability of as many as m bases being paired is given by the cumulative binomial distribution:

$$P(M \geq m) = \sum_{M=m}^{n} \frac{n!}{m!(n-m)!} f^m (1-f)^{n-m}$$

As a specific example, of the 18 bases involved in codons for the six invariant amino acids of polymerase sequence motifs (marked with asterisks in Fig. 2), 14 are paired bases. Substitution of the equation above with $n = 18$, $m = 14$, and $f = 0.57$ gives $P = 0.058$. This P value indicates that if 18 bases were randomly selected from the predicted RT structure, there would be only about a 6% chance of finding as many as 14 involved in base pairs.

RESULTS AND DISCUSSION

Computed RNA structures, such as those derived in the present study for the RT-coding region, probably only represent snapshots of a more dynamic structural picture involving fluctuations between various local minima and do not take into account the tertiary structures produced by stabilizing interactions between bases in different loops. Nonetheless, the most stable stem-loop structures in a computed minimum energy structure are likely to contribute to the preferred conformations of an actual RNA structure. Zuker et al. (44) showed that the single best predicted structure (relative to actual secondary structures) can be found within a set of related low-energy structures, usually with a free energy value within 2% of the computed global minimum, and that structures that are about 80% correct can be expected.

For the significance and stability of the predicted RNA structures, the following results (expressed as the number of SD below the average for the randomized sequences) were obtained for the various RT domains: (i) fingers, $z = 1.2$ SD, $P < 0.12$; (ii) palm, $z = 2.0$ SD, $P < 0.03$; (iii) thumb, $z = 4.1$ SD, $P < 0.0001$; and (iv) entire RT sequence used in Fig. 1, $z = 2.1$ SD, $P < 0.02$. The calculated free energies of folding (in kilocalories per mole) were -32.2, -84.2, and -60.8 for the fingers, palm, and thumb domains, respectively.

The low SD statistic for the fingers domain arises in part because of the large loops predicted for this region; these unpaired bases do not contribute to the energy but increase the probability of base pairings in the randomized sequences. In contrast, the palm and thumb domains and the RT-coding region taken as a whole have significant ($P < 0.03$) to highly significant ($P < 10^{-4}$) energies relative to those of the randomized sequences. These motif-containing domains (palm and thumb) are also generally more highly conserved in RTs than the N-terminal fingers region.

This result suggests that stable structures at the RNA level, even in coding regions, probably confer some evolutionary advantage and must be selected for by some mechanism, since the predicted structures overall are significantly more stable than what would be expected for random nucleotide sequences

TABLE 1. Mutations in HIV-1 RT associated with drug resistance

Drug[a]	Codon no.	Base no.	Base for wild type	Base mutation	Amino acid for wild type	Amino acid mutation	Reference
AZT	41	121	ATG	CTG	Met	Leu	19
	67	199	GAC	AAC	Asp	Asn	19
	70	209	AAA	AGA	Lys	Arg	19
	215	643, 644	ACC	TAC or TTC	Thr	Tyr/Phe	19
	219	655	AAA	CAA	Lys	Gln	19
	125	374	TTG	TGG	Leu	Trp	16
	142	424	ATT	GTT	Ile	Val	16
	294	880	CCA	ACA	Pro	Thr	16
	60	178	GTA	ATA	Val	Ile	36
	138	413	GAG	GCG	Glu	Ala	26
	210	629	TGG	TTG	Trp	Leu	26
	248	742	GAA	AAA	Glu	Lys	26
	259	776	AAG	AGG	Lys	Arg	26
DDI	74	220	TTA	GTA	Leu	Val	39
	135	403	ATA	GTA	Ile	Val	14
DDC	65	194	AAA	AGA	Lys	Asn	6
	69	205, 206	ACT	GAT	Thr	Asp	10
	165	494	ACA	ATA	Thr	Ile	10
DDG[b]	89	266	GAA	GGA	Glu	Gly	31
(−)-FTC or 3TC	184	550†, 552	ATG	GTG[c] or GTA	Met	Val	11, 37, 41
			ATG	ATA	Met	Ile	11, 37, 41
AZT + DDI	108	323	GTA	GCA	Val	Ala	12
	135	404	ATA	ACA	Ile	Thr	12
	50	148	ATT	GTT	Ile	Val	12
DDI + DDC	184	550	ATG	GTG	Met	Val	13
	294	880	CCA	TCA	Pro	Ser	13
AZT + nevirapine	106	317	GTA	GCA	Val	Ala	18
Pyridinones	98	293	GCA	GGA	Ala	Gly	5
	101	301	AAA	GAA	Lys	Glu	5
	103	309	AAA	AAC	Lys	Asn	5, 28
	108	322	GTA	ATA	Val	Ile	5
	179	536	GTT	GAT	Val	Asp	5
	179	537	GAT	GAPu	Val	Glu	5
Nevirapine and pyridinones	181	542	TAT	TGT	Tyr	Cys	19, 23, 24, 34
TIBO	100	298	TTA	ATA	Leu	Ile	24
	188	562	TAT	CAT	Tyr	His	1
Cl-TIBO or E-EPUSe	188	563	TAT	TGT	Tyr	Cys	1, 27
TSAO-m^{1}Hx or TSAO-T	138	412	GAG	AAG	Glu	Lys	2
BHAP	236	707	CCT	CTT	Pro	Leu	9
	181, 188	542, 563	TAT	TGT	Tyr	Cys	9
	238	713	AAA	ACA	Lys	Thr	32
PFA	161	482	CAA	CTA	Gln	Leu	25
	208	622	CAT	TAT	His	Tyr	25

[a] Abbreviations: AZT, 3′-azido-3′-deoxythymidine; BHAP, bis(heteroaryl)piperazine derivative known as atevirdine; Cl-TIBO, 9-chloro-TIBO; TSAO-m^{1}Hx, [2′,5′-bis-*O*-(*tert*-butyldimethylsilyl)-β-D-ribofuranosyl]-3′-spiro-5″-(4″-amino-1″,2″-oxathiole-2″,2″-dioxide)-N^9-hypoxanthine; DDC, 2′,3′-dideoxycytidine; DDG, 2′,3′-dideoxyguanosine; E-EPUSe, 1-(ethoxymethyl)-5-ethyl-6-(phenylselenenyl)uracil; (−)-FTC, (−)-β-2′,3′-dideoxy-5-fluoro-3′-thiacytidine; PFA, phosphonoformate; 3TC, (−)-β-2′,3′-dideoxy-3′-thiacytidine; TIBO, (+)-(5*S*)-4,5,6,7-tetrahydro-5-methyl-6-(3-methyl-2-butenyl)imidazo[4,5,1-*jk*][1,4]benzodiazepin-2(1*H*)-thione; TSAO-T, TSAO-thymine.

[b] Note that whereas RT was resistant to 2′,3′-dideoxyguanosine triphosphate, the virus in cell culture was not.

[c] Most commonly found in primary cell cultures and in humans infected with HIV-1 and treated with (−)-β-2′,3′-dideoxy-3′-thiacytidine.

of the same size and base composition. However, there is probably considerable variation in the details of the corresponding real RNA structures, even between strains of related retroviruses. We have repeated this type of analysis for other RNA viruses and found statistically significant predicted structures in regions coding for viral proteins. For example, for all the major proteins of hepatitis A virus, predicted RNA structures with computed free energies averaging about 3.4 SD ($P < 0.001$) below the average for the randomized sequences have been found (39a).

Our results for the HIV RT-coding region are also consistent with recently published results by Le et al. (21). Using their own RNA structure prediction method, they examined the entire HIV-2 genome for significant and stable RNA secondary structures. They list the *pol* gene region as being among the most significant and stable RNA structure region in HIV-2, but did not show any predicted structures in their report.

On the basis of the predicted structures, we can now analyze the distributions of DRMs and bases involved in codons for conserved amino acids in the coding region of the HIV-1 RT gene using the binomial distribution. Examination of Fig. 1 and 2 suggests that the DRMs are consistently associated with plateau regions (Fig. 1), which correspond to loops, bulges, or bends in the RNA secondary structure (Fig. 2). The four DRMs (10%) that do not fall precisely on loops, boundary bases, or G-U base pairs (base 220, 298, 542, and 563; Table 1) are nonetheless only a single base away from one. Although

base 707 does not appear to be on a plateau in Fig. 1, it falls on a border base pair in a concave bend in the structure (Fig. 2C); thus, we did not count it as an exception. Since mutation 537 occurs only in combination with the loop-expanding mutation 536, it was classified as a border base. The rarely observed mutation 552 (compared with mutation 550 for the same codon) is on a weak G-U base pair, which the mutation converts to a more stable A-U base pair; similarly, mutation 323 converts a G-U pair to a more stable G-C pair.

Analysis of the distribution of DRMs shows that the mutations are clustered near the ends of stem-loop structures (tops of mountains in Fig. 1) and at the juncture between loops or bulges and adjacent helical (stem) regions, i.e., near the bases of loops or at base pairs that form boundaries between stem and bulge regions (border bases). This clustering is statistically significant; 12 of the 42 DRMs fall precisely on such mountaintop plateau regions, which make up only 14.5% of the RT genome. Assuming that there is no relationship between RNA structure and DRM location (the null hypothesis), the probability of obtaining the observed distribution (i.e., as many as 12 of 42) by chance is $P < 0.015$ when P (mountaintop plateau) = 14.5% (i.e., $f = 0.145$). If this group is expanded to include all bases no more than 3 base pairs from the tops of the mountains, a set which makes up 19% of all bases ($f = 0.19$), 18 of the 42 DRMs fall in this set ($P < 0.0004$). This demonstrates a highly significant clustering of DRMs near the ends of the stem-loop structures. Similarly, the overall probability of a paired base being a border base is 52% for the entire RT-coding region, yet of the DRMs that fall on paired bases, 18 of 24 are border bases ($P < 0.02$). On the basis of current data, the association between the comparatively rare G-U base pairs (5% of all bases) and DRMs (4 of 42 bases; $P < 0.16$) must remain hypothetical. However, on the basis of chemical properties, it is clearly justifiable to categorize such G-U base pairs as unstable and group them with the unpaired and border bases.

This clustering of DRMs near the bases of loops or at border base pairs that form the boundary between helical stems and bulge regions suggests that the RNA structures involved may provide a physical constraint that interferes with the ability of the template to occupy the polymerase active site, enhancing the probability of errors. This interpretation is consistent with some previous work that has demonstrated that the accuracy and processivity of RT are affected by the sequence and secondary structure of the template (7, 33). Alternatively, bases on or near plateau regions are more likely to be unpaired than bases involved in stems, and are thus more likely to be exposed to chemical factors in the cellular environment that may enhance covalent modification or tautomerization of bases, leading to incorrect base pairing and misincorporation. A similar mechanism was proposed by Le et al. (20) to explain their observation of a correlation between hypervariable (i.e., mutation-prone) regions of the HIV *env* protein and what they called "open," or nonhelical, regions of the RNA structure predicted for the *env*-coding region. Their results clearly parallel our findings for the RT region.

The mapping of regions coding for the key amino acids of polymerase sequence motifs onto the helical nonplateau regions is also highly significant. These regions, making up 25% of the bases in the RT genome, consist of paired bases that are stacked between other paired bases, forming extended helical regions (stable paired bases), and do not include bases involved in G-U pairs, or boundary bases, which appear to be mutation-prone. Codons for the six invariant amino acids (30) associated with RNA-dependent polymerase motifs A to E (indicated by asterisks in Fig. 2) are associated with an exceptionally high ratio of stable bases (11 of 18); the probability of this occurring by chance is $P < 0.0015$ when $f = 0.25$. The apparent exception is the conserved lysine (K) of motif D (Fig. 2B); however, this codon is part of a poly(A) loop that may form base pairs with a UUU loop on the motif C stem to stabilize a three-dimensional folded structure. If larger segments of the polymerase motif regions are considered, including adjacent residues well conserved in HIV and related retroviruses (DVGD, LPQG, YMDD, KHQ, and MGY for motifs A to E, respectively), 28 of 54 bases are stable paired bases ($P < 10^{-5}$) and 43 of 54 bases are paired, in comparison with only 57% paired bases in the entire RT-coding region ($P < 0.0005$). Consistent with this correlation, the three longest uninterrupted helical stretches predicted for the RT-coding region fall precisely on motifs E, C, and A (in order of decreasing length). Significantly, only 4 of the 42 DRMs fall on a stable base ($P < 0.015$); in each case, however, these are adjacent to a border base pair and, in one case (mutation 542), is adjacent to both a border base pair and a weak G-U base pair (Fig. 2).

This correspondence between sequence motifs and helical regions suggests a possible evolutionary role and selection mechanism for these RNA structures. The motif residues are critical for enzyme function, and thus viral survival. Therefore, variants in which their codons are protected from mutation would by necessity be selected. This would also ensure greater replicative efficiency, since less nonviable viral RNA would be synthesized.

Finally, it must be noted that the predicted RNA secondary structure for the RT-coding region (Fig. 1) is also consistent with the possibility that the RT gene may be segmented at the RNA level into discrete structures corresponding to the functional domains of the protein. The predicted structures shown in Fig. 1 and 2 are obtained whether one uses the entire RT-coding region or performs three separate FOLD runs on the isolated finger, palm, and thumb domains. While this may be coincidental, it could be of considerable evolutionary significance. Segmentation of viral genomes by the formation of stable internally H-bonded nucleic acid structures corresponding to functional protein modules would provide a basis for the equivalent of exon shuffling in a single-stranded genome, even in the absence of introns. This in essence would be exons without introns; since exons are only defined by the locations of introns, this suggests that we need a new term for an RNA structural domain encoding a distinct protein domain. Precisely this type of segmentation of RNA genomes, by means of folding in coding regions, was recently proposed as a hypothetical mechanism for the insertion of introns in preferred locations in genes and as a means of enhancing genetic stability in RNA genomes (43). The present study is the first attempt to demonstrate that such structures may still exist in RNA viruses and may have features that can be correlated with experimental data, in this case, the locations of DRMs.

There is substantial evidence in the literature that secondary structure in coding regions is both present and potentially significant. For example, data on the effects of precise missense and silent mutations in poliovirus have been interpreted as evidence that the RNA secondary structure in coding regions plays a role in viral infection (4). It has recently been suggested that "the overall (RNA) structure of the HIV-2 genome has evolved to facilitate an optimal interaction with its t-RNA primer" (3). There is also direct structural evidence that, even in eukaryotic mRNA, there are stable structural features in coding regions as well as in introns (40). However, it must be acknowledged that, on the basis of a survey of the current literature, the question of RNA structure in coding regions is

largely unexplored. However, the recent work on the HIV *env*-coding region (20) clearly sets a precedent for precisely the sort of correlation between RNA structure and mutational liability that we have observed for the RT-coding region.

We have presented evidence that there is a statistically significant correlation by various criteria between the location of DRMs and the RNA structure of the HIV-1 RT-coding region. The mapping of known DRMs onto plateau regions of the predicted RNA secondary structures argues persuasively that RNA secondary structure in the template (in this case, the RT gene itself) is a previously unrecognized factor contributing to base misincorporation errors during reverse transcription. However, this may occur by an indirect mechanism, such as covalent modification of unprotected bases prior to reverse transcription. The highly significant mapping of codons for the key conserved residues of polymerase sequence motifs onto the apparently less mutation-prone extended helical regions suggests an obvious evolutionary rationale for both the existence and function of the predicted RNA structures and is consistent with a recent proposal that RNA structure in coding regions could enhance genetic stability (43).

Since, by necessity, our hypothesis has been based on predicted RNA secondary structures for the RT-coding region, its definitive testing must await the availability of experimental data on the actual structure of the viral RNA. However, the fact that we obtained such a striking correlation with both the mutation data and the location of motif codons suggests that the predicted structures are partially correct or, at the least, provide a meaningful measure of some thermodynamic property of the RNA chain that influences sites of mutation. It also must be noted that, combined with a precise knowledge of the RT protein structure (17), the ability to predict which amino acids are least likely to mutate could theoretically be used to design RT inhibitors less vulnerable to the problem of drug resistance.

ACKNOWLEDGMENTS

We are indebted to Z. Lesnikowski (Emory University) for helpful comments and M. Weise (University of Georgia) for assistance in automating the randomized RNA folding routine.

This work was supported in part by Public Health Service grants AI-30392 (to E.W.T.) and AI-25899 (to R.F.S.) from the National Institute of Allergy and Infectious Diseases, by the U.S. Department of Veterans Affairs, and by the Georgia Veterans Affairs Research Center for AIDS and HIV Infection.

REFERENCES

1. **Balzarini, J., A. Karlsson, M.-J. Pérez-Pérez, L. Vrang, J. Walbers, H. Zhang, B. Öberg, A.-M. Vandamme, M.-J. Camarasa, and E. De Clercq.** 1993. HIV-1 specific reverse transcriptase inhibitors show differential activity against HIV-1 mutant strains containing different amino acid substitutions in the reverse transcriptase. Virology **192**:246–253.
2. **Balzarini, J., S. Velazquez, A. San-Felix, A. Karlsson, M.-J. Pérez-Pérez, M.-J. Camarasa, and E. De Clercq.** 1993. Human immunodeficiency virus type-1 specific purine analogues show a resistance spectrum that is different from that of the human immunodeficiency virus type-1-specific non-nucleoside analogues. Mol. Pharmacol. **43**:109–114.
3. **Berkhout, B., and I. Schoneveld.** 1993. Secondary structure of the HIV-2 leader RNA comprising the tRNA-primer binding site. Nucleic Acids Res. **21**:1171–1178.
4. **Borzakian, S., I. Pelletier, V. Calvez, and F. Colbere-Garapin.** 1993. Precise missense and silent point mutations are fixed in the genomes of poliovirus mutants from persistently infected cells. J. Virol. **67**:2914–2917.
5. **Byrnes, V. W., V. V. Sardana, W. A. Schleif, J. H. Condra, J. A. Waterbury, J. A. Wolfgang, W. J. Long, C. L. Schneider, A. J. Schlabach, B. S. Wolanski, D. J. Graham, L. Gotlib, A. Rhodes, D. L. Titus, E. Roth, O. M. Blahy, J. C. Quintero, S. Staszewski, and E. A. Emini.** 1993. Comprehensive mutant enzyme and viral variant assessment of human immunodeficiency virus type 1 reverse transcriptase resistance to nonnucleoside inhibitors. Antimicrob. Agents Chemother. **37**:1576–1579.
6. **D'Aquila, R. T., A. M. Caliendo, D. Zhang, et al.** 1993. Zalcitabine and didanosine resistance mutations have different effects on in vitro RT function than zidovudine resistance mutations. Second International HIV-1 Drug Resistance Workshop, Noordwijk, The Netherlands.
7. **DeStefano, J. J., R. G. Buiser, L. M. Mallaber, P. J. Fay, and R. A. Bambara.** 1992. Parameters that influence processive synthesis and site-specific termination by human immunodeficiency virus reverse transcriptase on RNA and DNA templates. Biochim. Biophys. Acta **1131**:270–280.
8. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
9. **Dueweke, T. J., T. Pushkarskaya, S. M. Poppe, S. M. Swaney, J. Q. Zhao, I. S. Y. Chen, M. Stevenson, and W. G. Tarpley.** 1993. A mutation in reverse transcriptase of bis(heteroaryl)piperazine-resistant human immunodeficiency virus type 1 that confers increased sensitivity to other nonnucleoside inhibitors. Proc. Natl. Acad. Sci. USA **90**:4713–4717.
10. **Fitzgibbon, J. E., R. M. Howell, C. A. Haberzettl, S. J. Sperber, D. J. Gocke, and D. T. Dublin.** 1992. Human immunodeficiency virus type 1 *pol* gene mutations which caused decreased susceptibility to 2′,3′-dideoxycytidine. Antimicrob. Agents Chemother. **36**:153–157.
11. **Gao, Q., Z. Gu, M. A. Parniak, J. Cameron, N. Cammack, C. Boucher, and M. A. Wainberg.** 1993. The same mutation that encodes low-level human immunodeficiency virus type 1 resistance to 2′,3′-dideoxyinosine and 2′,3′-dideoxycytidine confers high-level resistance to the (−) enantiomer of 2′,3′-dideoxy-3′-thiacytidine. Antimicrob. Agents Chemother. **37**:1390–1392.
12. **Gao, Q., Z. Gu, M. N. Parniak, X. Li, and M. A. Wainberg.** 1992. In vitro selection of variants of human immunodeficiency virus type 1 resistant to 3′-azido-3′-deoxythymidine and 2′,3′-dideoxyinosine. J. Virol. **66**:12–19.
13. **Gu, Z., Q. Gao, X. Li, M. N. Parniak, and M. A. Wainberg.** 1992. Novel mutation in the human immunodeficiency virus type 1 reverse transcriptase gene that encodes cross-resistance to 2′,3′-dideoxycytidine. J. Virol. **66**:7128–7135.
14. **Gu, Z., and M. A. Wainberg.** Personal communication.
15. **Hogeweg, P., and B. Hesper.** 1984. Energy directed folding of RNA sequences. Nucleic Acids Res. **12**:67–74.
16. **Japour, A. J., P. A. Chatis, H. A. Eigenrauch, and C. S. Crumpacker.** 1991. Detection of human immunodeficiency virus type 1 clinical isolates with reduced sensitivity to zidovudine and dideoxyinosine by RNA.RNA hybridization. Proc. Natl. Acad. Sci. USA **88**:3092–3096.
17. **Kohlstaedt, L. A., J. Wang, J. M. Friedman, P. A. Rice, and T. A. Steitz.** 1992. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science **256**:1783–1790.
18. **Larder, B. A.** 1992. 3′-Azido-3′-deoxythymidine resistance suppressed by a mutation conferring human immunodeficiency virus type 1 resistance to nonnucleoside reverse transcriptase inhibitors. Antimicrob. Agents Chemother. **36**:2664–2669.
19. **Larder, B. A., and S. D. Kemp.** 1989. Multiple mutations in HIV-1 reverse transcriptase confer high level resistance to zidovudine (AZT). Science **246**:1155–1157.
20. **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 isolates. Nucleic Acids Res. **17**:3275–3288.
21. **Le, S.-Y., J.-H. Chen, and J. V. Maizel.** 1991. Detection of unusual RNA folding regions in HIV and SIV sequences. CABIOS **7**:51–55.
22. **Le, S.-Y., and J. V. Maizel, Jr.** 1989. A method for assessing the statistical significance of RNA folding. J. Theor. Biol. **138**:495–510.

23. **Mellors, J. W., G. E. Dutschman, G.-J. Im, E. Tramontano, S. R. Winkler, and Y.-C. Cheng.** 1992. *In vitro* selection and molecular characterization of human immunodeficiency virus-1 resistant to non-nucleoside inhibitors of reverse transcriptase. Mol. Pharmacol. **41**:446–451.
24. **Mellors, J. W., G. J. Im, E. Tramontano, S. R. Winkler, D. J. Medina, G. E. Dutschman, H. Z. Bazmi, G. Piras, C. J. Gonzalez, and Y.-C. Cheng.** 1993. A single conservative amino acid substitution in the reverse transcriptase of human immunodeficiency virus virus-1 confers resistance to (+)-(5S)-4,5,6,7-tetrahydro-5-methyl-6-(3-methyl-2-butenyl)imidazo [4,5,1-jk][1,4]benzodiazepin-2(1H)-thione (TIBO R82150). Mol. Pharmacol. **43**:11–16.
25. **Mellors, J. W., and R. F. Schinazi.** Unpublished data.
26. **Muckenthaler, M., N. Gunkel, P. Levantis, K. Broadhurst, B. Goh, B. Colvin, G. Forster, G. G. Jackson, and J. S. Oxford.** 1992. Sequence analysis of an HIV-1 isolate which displays unusually high-level AZT resistance *in vitro*. J. Med. Virol. **36**:79–83.
27. **Nguyen, M. H., R. F. Schinazi, N. Goudgaon, H. Z. Bazmi, and J. W. Mellors.** 1993. HIV-1 resistance to acyclic 6-substituted pyrimidines. Program Abstr. 33rd Intersci. Confer. Antimicrob. Agents Chemother., abstr. 678.
28. **Nunberg, J. H., W. A. Schleif, E. J. Boots, J. A. O'Brien, J. C. Quintero, J. M. Hoffman, E. A. Emini, and M. E. Goldman.** 1991. Viral resistance to human immunodeficiency virus type 1-specific pyridinone reverse transcriptase inhibitors. J. Virol. **65**:4887–4892.
29. **Osterburg, G., and R. Sommer.** 1981. Computer support of DNA sequence analysis. Computer Programs Biomed. **13**:101–109.
30. **Poch, O., I. Sauvaget, M. Delarue, and N. Tordo.** 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. **8**:3867–3874.
31. **Prasad, V. R., I. Lowy, T. de los Santos, L. Chiang, and S. P. Goff.** 1991. Isolation and characterization of a dideoxyguanosine triphosphate-resistant mutant of human immunodeficiency virus reverse transcriptase. Proc. Natl. Acad. Sci. USA **88**:11363–11367.
32. **Resnick, L.** Personal communication.
33. **Ricchetti, M., and H. Buc.** 1990. Reverse transcriptases and genomic variability: the accuracy of DNA replication is enzyme specific and sequence dependent. EMBO J. **9**:1583–1593.
34. **Richman, D., C.-K. Shih, I. Lowy, J. Rose, P. Prodanovich, S. Goff, and J. Griffin.** 1991. Human immunodeficiency virus type 1 mutants resistant to nonnucleoside inhibitors of reverse transcriptase arise in tissue culture. Proc. Natl. Acad. Sci. USA **88**:11241–11245.
35. **Richman, D. D.** 1993. Resistance of clinical isolates of human immunodeficiency virus to antiretroviral agents. Antimicrob. Agents Chemother. **37**:1207–1213.
36. **Rubsamen-Waigmann, H., B. Schroder, L. Biesert, C. D. Bauermeister, H. Von-Briesen, H. Suhartono, F. Zimmermann, H. D. Brede, A. Regeniter, and S. Gerte.** 1991. Markers for HIV-disease progression in untreated patients and patients receiving 2-microglobulin, $CD4^+$ cell counts, and HIV antigen. Infection **19**:S77–S82.
37. **Schinazi, R. F., R. M. Lloyd, Jr., M. Nguyen, D. L. Cannon, A. McMillan, N. Ilksoy, C. K. Chu, D. C. Liotta, H. Z. Bazmi, and J. W. Mellors.** 1993. Characterization of human immunodeficiency viruses resistant to oxathiolane-cytosine nucleosides. Antimicrob. Agents Chemother. **37**:875–881.
38. **Schinazi, R. F., J. R. Mead, and P. M. Feorino.** 1992. Insights into HIV chemotherapy. AIDS Res. Hum. Retroviruses **8**:963–990.
39. **St. Clair, M. H., J. L. Martin, G. Tudor Williams, M. C. Bach, C. L. Vavro, D. M. King, P. Kellam, S. D. Kemp, and B. A. Larder.** 1991. Resistance to ddI and sensitivity to AZT induced by a mutation in HIV-1 reverse transcriptase. Science **253**:1557–1559.

39a. **Taylor, E. W., and R. K. Jalluri.** Unpublished data.

40. **Teare, J., and P. Wollenzein.** 1990. The structure of pre-mRNA molecule in solution determined with a site directed cross-linking reagent. Nucleic Acids Res. **18**:855–864.
41. **Tisdale, M., S. D. Kemp, N. R. Parry, and B. A. Larder.** 1993. Rapid *in vitro* selection of human immunodeficiency virus type 1 resistant to 3′-thiacytidine inhibitors due to a mutation in the YMDD region of reverse transcriptase. Proc. Natl. Acad. Sci. USA **90**:5653–5656.
42. **Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Alizon.** 1985. Nucleotide sequences of the AIDS virus, LAV. Cell **40**:9–17.
43. **Wistow, G.** 1993. Protein structure and introns. Nature (London) **364**:107–108.
44. **Zuker, M., J. A. Jaeger, and D. H. Turner.** 1991. A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures determined by phylogenetic comparison. Nucleic Acids Res. **19**:2707–2714.
45. **Zuker, M., and P. Steigler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133–148.