JOURNAL OF EXPERIMENTAL ZOOLOGY (MOL DEV EVOL) 285:119–127 (1999)

Viral RNA and Evolved Mutational Robustness

ANDREAS WAGNER[1,3]* AND PETER F. STADLER[2,3]
[1]*Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131-1091*
[2]*Institut für Theoretische Chemie, Universität Wien, Währingerstraße 17, A-1090 Wien, Austria*
[3]*The Santa Fe Institute, Santa Fe, New Mexico 87501*

ABSTRACT Many properties of organisms show great robustness against mutations. Whether this robustness is an evolved property or intrinsic to genetic systems is by and large unknown. An evolutionary origin of robustness would require a rethinking of key concepts in the field of molecular evolution, such as gene-specific neutral mutation rates, or the context-independence of deleterious mutations. We provide evidence that mutational robustness of the genome of RNA viruses to mutational changes in secondary structure has evolved. *J. Exp. Zool.* (*Mol. Dev. Evol.*) *285:119–127, 1999.* © 1999 Wiley-Liss, Inc.

Even the simplest known organisms consist of thousands of molecular components which form a highly integrated network of interactions. An immense number of parameters that specify these interactions have to be finely tuned to ensure survival and reproduction of a cell. With this view in mind, and considering that organisms are constantly subject to mutations, it may seem surprising that multicellular organisms evolved at all. When layer upon layer of suborganismal organization is accumulated in evolution, surely there must be a point at which a system's complexity becomes unsustainable, at which the "house of cards" collapses?

In contrast to this perspective, many subsystems of organisms are quite robust to mutations, as exemplified by the high tolerance of proteins to amino acid changes (Bowie et al., '90), the resilience of metabolic flux to changes in enzyme activity (Kacser et al., '81; Hartl et al., '85), and the robustness of developmental pathways to mutations in their constituent genes (Dun and Fraser, '58; Rendel, '79; Joyner et al., '91; Tautz, '92; Weintraub, '93; Cadigan et al., '94). The mechanistic basis for robustness in each of these cases depends on the peculiarities of the system studied, raising the possibility that robustness is an intrinsic characteristic of these systems. A less explored alternative is that observed robustness is a property that has evolved, perhaps because it protects a biological system against deleterious mutations. Here we show that evolution of robustness took place in the functional RNA secondary structures of RNA viral genomes.

In contrast to conventional scenarios of natural selection, where selection acts directly on fitness differences in genetic variants, the selective advantage of a robust system is its decreased susceptibility to deleterious mutations. Because of this indirect, "second order," and thus slow mechanism it is highly unlikely that the evolution of robustness can be directly observed in the laboratory (Wagner, '99). However, the outcome of this process can be observed: Features of biological systems that are under very strong stabilizing selection, and thus highly conserved, should display higher mutational robustness than non-conserved ones. (It must be noted that an increase in robustness need not necessarily increase population mean fitness. This is because mean equilibrium fitness does not depend on the average effect of a mutation, but only on the mutation rate, a population genetic principle often referred to as the Haldane-Muller principle.) Anecdotal evidence that can be interpreted in this way derives from a conspicuous lack of variation in many basic Bauplan characters, and has led to the concept of genetic canalization of these characters, implying robustness in the developmental pathways generating them (Waddington, '57; Rendel, '79). However, with the exception of a study on life history

Grant sponsor: Santa Fe Institute; Grant sponsor: Austrian Fond zur Förderung der Wissenschaftlichen Forschung; Grant number: P12591-INF.
*Correspondence to: Andreas Wagner, Dept. of Biology, University of New Mexico, 167A Castetter Hall, Albuquerque, NM 87131-1091. E-mail: wagnera@unm.edu
Received 15 February 1999; Accepted 4 March 1999.

traits in *Drosophila* (Stearns and Kawecki, '94), to our knowledge no systematic effort has been made to test this prediction.

The genome of RNA viruses presents us with an ideal study system in this regard. Parts of the genomic RNA form conserved secondary structures that are crucial for regulating the viral life cycle, e.g., the TAR (Baudin et al., '93) and the RRE (Dayton et al., '92) elements of HIV1, the 3′ stem-loop structures of Flaviviruses (Brinton et al., '86), the IRES region of picorna viruses (Jackson and Kaminski, '95), or the 3′UTR of flaviviruses (Mandl et al., '98). A large fraction of the genome, on the other hand, encodes proteins. These regions form secondary structures of little or no functional importance but with thermodynamic stabilities comparable to functionally important secondary structures. A single RNA virus genome thus contains both conserved structural elements, which we expect to display high mutational robustness, and non-conserved elements which serve as the non-robust control.

METHODS

The number of experimentally documented conserved secondary structure elements is too small to yield anything but anecdotal evidence for our proposition (Huynen et al., '93). A computational procedure, however, has recently become available that reliably detects conserved secondary structure motifs from a moderate sample of related RNA sequences (Hofacker et al., '98; Hofacker and Stadler, '99) by combining thermodynamics-based secondary structure predictions and comparative sequence analysis. It extracts conserved secondary structure elements from moderately sized data sets of related viral sequences using a combination of thermodynamic secondary structure prediction and multiple sequence alignments. A related procedure is described in Lück et al. ('96). However, because the procedure introduced by Hofacker ('98) and Hofacker and Stadler ('99) uses additional information, namely the presence or absence of compensatory mutations, to detect conserved structural motifs, it is likely to produce fewer false positives.

RNA secondary structures are computed using a robust and accurate minimum energy folding algorithm (Zuker and Sankoff, '84), as implemented in the Vienna RNA Package, version 1.2 (Hofacker et al., '94). The energy parameters used by the Vienna RNA Package are based on Freier et al. ('86) ,Jaeger et al. ('89), and He et al. ('91). Complete RNA genomes were folded on CalTech's Delta using the message passing version of the minimum folding algorithm described in Hofacker et al. ('96a,b). The computed structures are in good agreement with experimentally determined secondary structures (Tacker et al., '96; Huynen et al., '97; Schuster et al., '97).

Dynamic programming-based RNA folding algorithm at present cannot predict pseudo-knots and other "tertiary" interactions such as G-quartets. If such elements are present, the prediction may deviate significantly from the true structure. This complication, however, has very little or no impact on the present study, because different sequences forming a conserved pseudo-knot will yield inconsistent predictions. Hence no conserved structure would be predicted, and such regions therefore would not enter our data set.

The software package Clustal W (Thompson et al., '94) is used to generate a multiple alignment of the viral sequences. The program alidot (Hofacker et al., '98) translates the sequence alignment into an alignment of the secondary structures. It does so by generating a list of all base pairs (i, j) occurring in the aligned minimum energy secondary structures of the entire viral data set. This list is then sorted according to the following hierarchical criteria: (1) The more sequences are non-compatible with (i, j), the less supported is the base pair. (2) Symmetric base pairs, i.e., base pairs (i, j) for which i is the most frequent pairing partner of j and j is the most frequently predicted pairing partner of i, are better supported than other base pairs. (3) A base pair with more consistent (compensatory) mutations is better supported. (4) Base pairs that are predicted more frequently are better supported. Scanning this sorted list of base pairs in the order of decreasing support, a base pair is removed if it conflicts with a higher-ranking one that has already been accepted. Finally, base pairs are collected into stems and those stems that are both predicted in a sizable fraction of the sequences and that are supported by sequence co-variations are retained in the final output. A detailed description of this procedure can be found in Hofacker et al. ('98).

The approach pursued here consists of a comparison of various measures of mutational stability between conserved and non-conserved viral secondary structure elements. Using the method just described, conserved secondary structure elements were extracted from three different groups of single-stranded RNA viruses, namely dengue virus (Genbank names: DEN2CGA, DEN2CMEMSA, DEN2CMEMSB, DEN2JAMCG, DEN2RCG, DEN3CME, DEN4STRA, DENT1SEQ), hepatitis

C virus (Genbank names: HCU16362, HCU45476, HPCCGAA, HPCCGENOM, HPCCGS, HPCEGS, HPCHCJ1, HPCJ483, HPCJRNA, HPCJTA, HPCK3A, HPCPP, HPCRNA; these sequences do not contain the 98nt long X-tail at the 3′ terminus), and HIV1 (Genbank names: HIVANT70, HIVBCSG3C, HIVCAM1, HIVD31, HIVELI, HIVLAI, HIVMAL, HIVMVP5180, HIVNDK, HIVOYI, HIVRF, HIVU455, HIVZ2Z6). The selected sequences represent a maximally diverse subset of the available completely sequenced genomes for each family. The mean pair-wise sequence similarities range from 77% to 81%. No conspicuous differences exist in sequence similarity among structurally conserved and non-conserved regions.

A total of 34, 9, and 13 conserved elements were isolated from the dengue, hepatitis C, and HIV groups, respectively. The data set contains elements of well-known functional significance such as the TAR-hairpin of HIV and the 3′ stem-loop structure of flavivirus RNA. For most of the structures, however, no function has been reported so far. Further, a sample of 181 non-conserved structural elements was isolated from representatives of each of the three groups (Jamaican genotype of dengue, GenBank acc. no. M20558, Japanese genotype of hepatitis C, acc. nos. D14484, D01173; African genotype of HIV, acc. no. L20587). This was done by (1) extracting randomly chosen fragments spanning 300 nucleotides and not overlapping a conserved element from each of these sequence data sets, (2) computing the secondary structures of these fragments, and (3) extracting, from each such fragment a sub-sequence corresponding to a well-defined secondary structure. A total of 64, 55, and 62 non-conserved elements was thus isolated from dengue, hepatitis C, and HIV, respectively.

For each of the extracted conserved and non-conserved elements, we determined two complementary measures of mutational stability. We first analyzed the effects of mutations inside an element by generating all possible single point mutations within all conserved and non-conserved elements ($3n$ mutations for an element of n nucleotides). For each point mutation we then numerically evaluated the minimum free energy secondary structure of the mutated element. We compared the secondary structure of a mutated element to that of the unmutated element using a variety of distance measures for secondary structures, including tree-edit distance and string distance (Hogeweg and Hesper, '84; Hofacker et al., '94). The qualitative results obtained were independent of the distance measure used.

Second, we determined the stability of secondary structure elements to changes in sequence context, i.e., to changes in sequences surrounding the element, in two ways: (1) We folded a window of the genomic consensus sequence containing both the conserved element and flanking sequences of different lengths, ranging from one nucleotide to the length of the element. We folded each of these windows, and compared the secondary structure of the segment encompassing the element of interest to its "wild-type" structure. (2) To emulate the effect of changes in sequence context, we attached random RNA sequences of varying length to a conserved element of interest, folded the resulting sequence, and compared the resulting structure to the "wild-type" as in (1). Ideally, one would of course wish to introduce mutations anywhere in the viral genome, and to evaluate their effect on any particular element by folding the entire mutated genome. This is computationally infeasible. Computing the secondary structures of long RNAs sequences is a time-consuming task since CPU requirements scale as $O(n^3)$ and memory requirements scale as $O(n^2)$ with the chain length n. The computation for a single HIV-1 sequence, $n \sim 9{,}200$, for instance, takes several hours on a massively parallel computer and requires more than 1 GByte of memory. These considerations motivated the more limited approach we pursued here.

RESULTS

The genetic perturbations that may affect the secondary structure of a short sequence element within a larger genome fall into two classes: those that take place within the element, and those that take place in the remainder of the genome. We compared the effect of point mutations in conserved and non-conserved elements by exhaustively mutating all positions within each isolated element in the way described above. Figure 1a and b shows extreme examples of a non-conserved and a conserved HIV-1 secondary structure element, respectively, along with histograms of the distribution of structure distances for all single point mutations. As judged by the average structure distance, the conserved element is six-fold more stable than the non-conserved element.

Figure 2a summarizes analogous results for all conserved and non-conserved structure elements, using two different structure distance measures. It shows that conserved elements are consistently more stable than non-conserved elements, although the stability increase of conserved elements is in

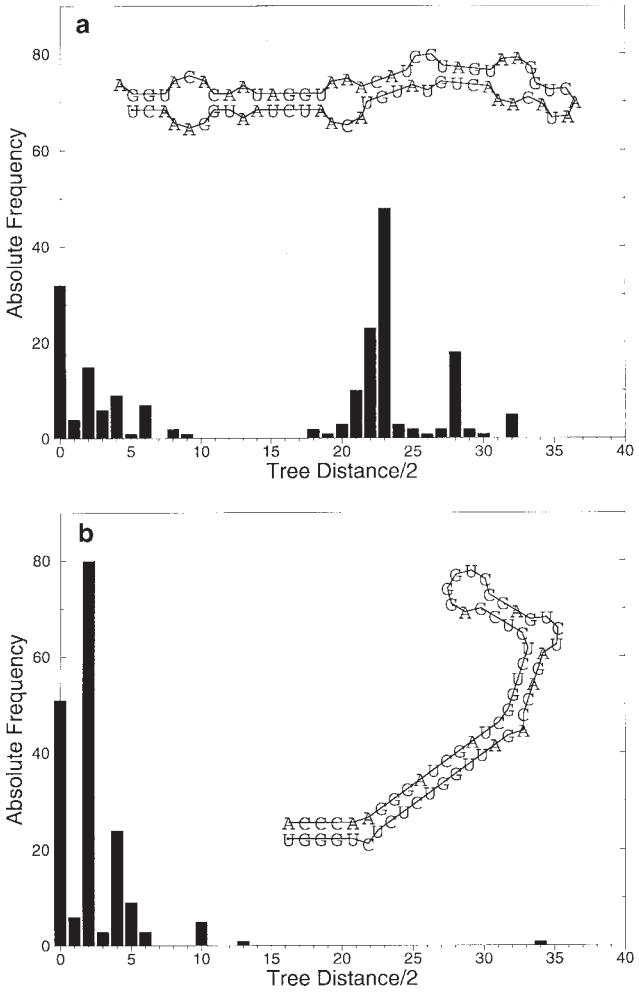

Fig. 1. Conserved and non-conserved elements differ in robustness. **a:** Absolute frequency of structure distances for all single point mutations of the *non-conserved* HIV-1 structural element shown in the inset (position 7089-7155 of the HIV-1 sequence HIVANT70, Genbank acc. no. M31171). The tree edit distance plotted here is one of several distance measures for RNA secondary structures (Fontana et al., '93b). Mean/standard error of structure distance normalized by length of element: $4.7 \times 10^{-1}/2.3 \times 10^{-2}$. **b:** Absolute frequency of structure distances for all single point mutations of a HIV-1 *conserved* secondary structure element. Mean/standard error of normalized distance: $7.7 \times 10^{-2}/7.8 \times 10^{-3}$.

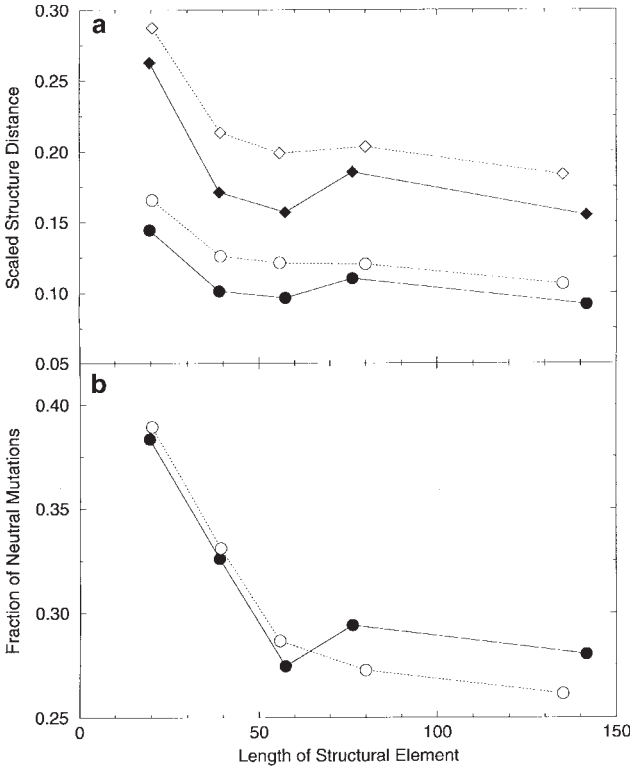

Fig. 2. Robustness against point mutations. **a:** Mean normalized tree edit distance (diamonds) and string distance (circles) for all single point mutations of non-conserved (open symbols) and conserved (black symbols) secondary structure elements, as a function of the primary sequence length. Elements were binned into length intervals (9–30 nucleotides: 19 conserved/73 non-conserved elements; 30–50: 19/45; 50–70: 9/28; 70–90: 3/15; 90–192: 6/20), for which the mean lengths are plotted on the abscissa. **b:** Mean fraction of point mutations that cause no deviation from the secondary structure of the native element for non-conserved (○) and conserved (●) structure elements. Differences in this fraction of *structurally* neutral mutations are only apparent for elements longer than 70 nucleotides.

general less than 50%. Figure 2b considers only the fraction of single point mutants whose secondary structure does not differ at all from that of the unmutated element. Here, conserved elements longer than 70 base pairs show a higher frequency of neutral mutations with respect to secondary structure than non-conserved elements.

The second class of mutations is located outside the secondary structure elements of interest. Due to the non-local nature of secondary structure formation they may nevertheless affect the sub-structure of interest. A distant position that underwent mutation may now interact preferably with the structural motif under consideration (in which case the structure of interest is disrupted), or with a position in its close vicinity, thereby changing the structures in its vicinity. Formerly unpaired regions may thus become part of a secondary structure motif. Conversely, sequences that were part of a helix before the mutation, may become available for interaction with the element of interest. Mutations that affect this "genomic context" of a sequence element are by no means rare (Fontana et al., '93b).

Because it is infeasible to compute the secondary structure of a complete genome for a large number of mutants, we assess the sensitivity of a structural element to changes in its genomic con-

text by adding either genomic or random flanking sequences of varying length as described in "Methods." Adding flanking genomic sequences of varying size emulates the effect of structural changes elsewhere in the molecule, whereas adding random sequences models the effect of changes in sequence context. The latter is motivated by the observation that many RNA viruses have regions with enormously high sequence variability. Insertions and deletions have an especially profound influence on this variability. Their abundance is exemplified by our sample of HIV-1 sequences, where approximately seven insertions or deletions per 1,000nt are observed. Another impressive example is the hypervariable part of the 3′ non-coding region in flaviviruses (Wallner et al., '95) which is located adjacent to a particularly well-conserved piece of secondary structure (Mandl et al., '98) but which itself has been completely randomized.

The effects of genomic and random context differ greatly: The gray scale in Figure 3a, e corresponds to the similarity in secondary structure: light (dark) points correspond to fragments with attained secondary structures that are similar (dissimilar) to the fragment in Figure 3d. String distance, that is, the Hamming distance of the dot-parenthesis encoding of the secondary structure (see Hogeweg and Hesper '84) was used as a structure distance measure. Despite the striking differences in the appearance of the two grid plots for genomic and random context, the overall distance (gray-level) increases in both as more nucleotides are added.

Figure 3b, c, f, g shows that conserved secondary structure elements display a higher degree of robustness to addition of genomic or random sequences than non-conserved elements, with the possible exception of sequences between 70 and 90 nucleotides. The number of sequences in this interval, however, is very small in our data set, and thus highly susceptible to statistical outliers.

Kendall's τ rank correlation coefficient between stability measures for 53 and 117 conserved and non-conserved structure elements, respectively, are listed in Table 1. Two different measures for stability to the sequence context (added 5′ and 3′ nucleotides) are considered, one in which stability to all contexts with less than 100 nucleotides is averaged, (context length $\leq$100, see also Fig. 3), and a second measure where the number of nucleotides added is equal to a constant fraction, $n/2$, of the length n of an element. (context length = $n/2$). All three stability measures, stability to point mutations, stability to varying genomic context lengths, and stability to random contexts, are significantly correlated. Significance tests were performed against the null-hypothesis that the correlation coefficients are equal to zero (Sokal and Rohlf, '81). All correlation coefficients are significant at $P < 0.05$. Pearson product-moment correlation coefficients (not shown) have values similar to those given here. There is a conspicuous increase in all three pairwise correlations from non-conserved to conserved fragments, although its significance can not be evaluated, given the non-normality of the underlying distributions.

In summary, our results indicate that conserved secondary structure elements from three groups of RNA viruses are more robust to genetic perturbations than their non-conserved counterparts, regardless of whether the perturbations occur inside or outside the element. This indicates that robustness in RNA secondary structures has evolved in RNA viruses.

DISCUSSION

Conventional methods of multiple sequence alignment can be used to identify highly conserved sequence elements. In analogy, a multiple alignment of secondary structures can be used to identify highly conserved structural elements. The sequence information, or more precisely, the presence or absence of compensatory mutations is utilized to verify the existence of a conserved element (Hofacker et al., '98). Thus structural elements are identified as conserved not because of their thermodynamic stability or because of sequence conservation but because there is a consistently predicted structure in the presence of sequence variation. This comparative approach compensates for the occasional inaccuracies in individually predicted secondary structures. The method has been applied to a variety of virus families and readily produces the known conserved elements such as TAR and RRE of HIV-1, the panhandle structure of Hanta virus, and the 3′UTR of flavivirus (Hofacker et al., '98; Hofacker and Stadler, '99).

Conservation of secondary structure features in sequences with about 80% pairwise identity (as in our sample) is likely to be a consequence of functional importance of the secondary structure. Statistical surveys (Fontana et al., '93a; Bonhoeffer et al., '93) show that even a small number of mutations is sufficient to alter secondary structure completely, and at a sequence divergence of only 10% the overwhelming majority of sequences will fold into structures with at most vague similarities. Conservation of secondary structure

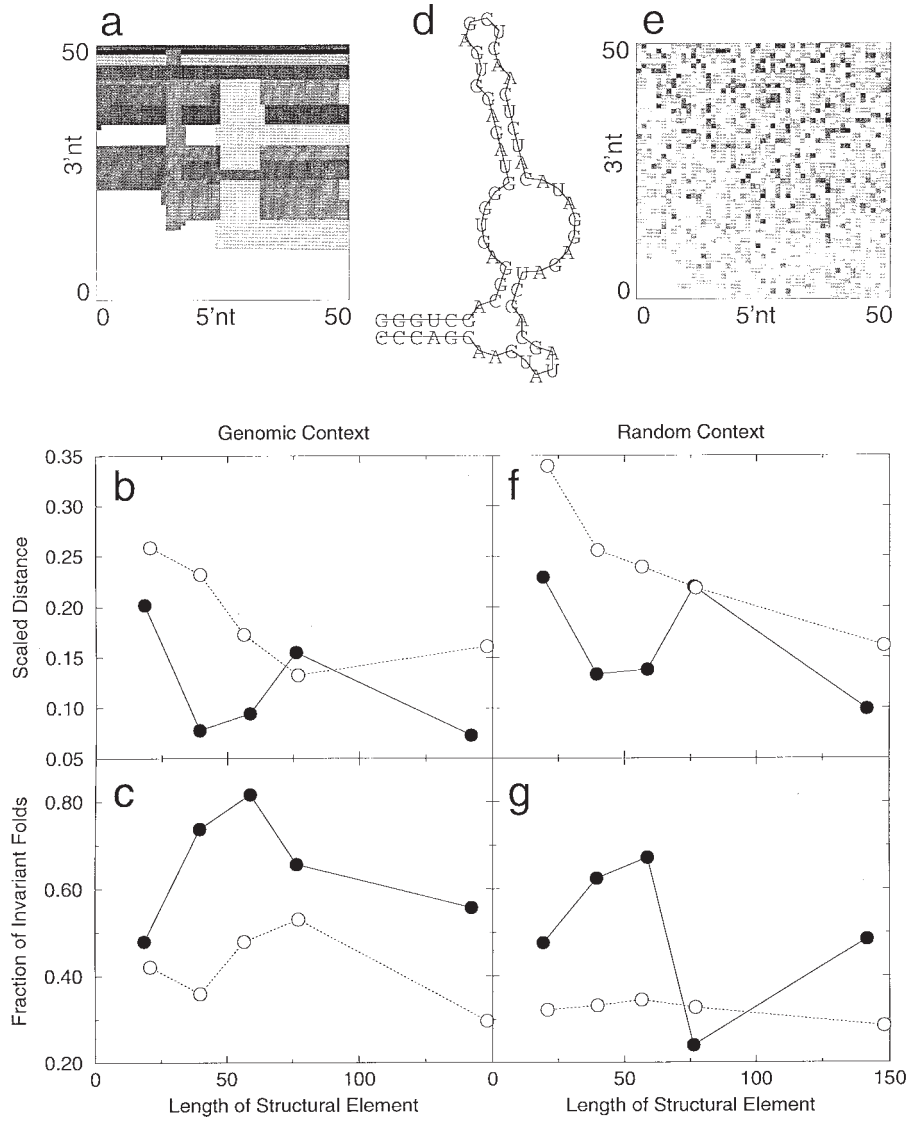

Fig. 3. Robustness of secondary structure elements to changes of the genomic context. We consider flanking sequences of different lengths at the 3′- and the 5′-end of a secondary structure element (d, taken from the dengue secondary structure alignment, positions 5399–5462) that are either part of the genomic consensus sequence, (a–c), or randomly generated sequences (e–g). The 2D grid plots (**a, e**) display the structure distance of the mutants to the original structure (**d**) as a function of the number of nucleotides added to the 5′ and 3′ ends (x- and y-axis, respectively). The mean structure distance (**b**) and the fraction of identical structures (**c**) for all fragments with up to 100 nucleotides of *genomic* sequence added to the fragment is plotted for 53 conserved (○) and 117 non-conserved (●) structure elements. Analogous data are shown in (**f, g**) for the addition of up to 50 *random* nucleotides.

among sequences with only 90% similarity thus should be seen as a consequence of stabilizing selection, indicating the functional importance of the structure rather than as a consequence of sequence similarity.

The importance of the in silico approach taken here is emphasized by the fact that the large numbers of perturbations analyzed here can not be easily studied in laboratory experiments (e.g., more than 20,000 point mutants were generated and folded). The computational approach, however, is not free of problems. While the similarity among the genome sequences used in this study is low enough to allow us to assume that conserved secondary structures have conserved function (Schuster et al., '94; Schuster and Stadler, '98), we do not know what this function is. Consequently, we do not know how much deviation in secondary structure is admissible for an element to be functional, and thus for a mutation to be

TABLE 1. Correlations (Kendall's τ rank correlation coefficient) between robustness measures for point mutations (p.m.), genomic context (g.c.) and random context (r.c.) for two different context lengths l.[1]

	l ≤ 100		l = n/2	
	τ	P	τ	P
Conserved Structured Elements				
p.m./g.c.	0.44	4×10^{-6}	0.31	1×10^{-3}
p.m./r.c.	0.59	6×10^{-10}	0.35	2×10^{-4}
g.c./r.c.	0.51	7×10^{-8}	0.46	1×10^{-6}
Non-Conserved Structure Elements				
p.m./g.c.	0.26	5×10^{-5}	0.15	2×10^{-2}
p.m./r.c.	0.57	$< 10^{-10}$	0.30	3×10^{-6}
r.c./r.c.	0.36	2×10^{-8}	0.41	2×10^{-10}

[1]See text for details.

functionally neutral, as opposed to structurally neutral. We take the difference in the measures of structural robustness as an indication for a similar difference in the fraction of functionally neutral mutations for conserved and non-conserved RNA sequences.

In theory, very simple differences among conserved and non-conserved sequences could be responsible for their differential mutational robustness. These might include differences in GC content and/or the number of paired bases. However, in our data set we did not find any obvious correlations between GC content and the number of paired bases on one hand, and mutational robustness on the other hand. Recently, Wuchty et al. ('99) reported that the "well definedness" of a secondary structure, that is, the stability of the minimum energy structure relative to the energetically most favorable suboptimal folds, is indeed correlated with mutational stability. Mutational stability of RNA secondary might therefore have arisen as the consequence of selection for well definedness or thermodynamic stability of the structural elements. However, many viral secondary structure motifs need to be sufficiently flexible to unfold for processes such as replication or translation. Furthermore, it has been shown experimentally that mutations leading to increased thermodynamic stability of several important viral secondary structure motifs have deleterious effects (Olsthoorn et al., '94; Honda et al., '96). While this evidence supports the hypothesis that mutational robustness has been selected for its own sake, a conclusive decision on this issue is not at hand, and would require a much larger set of experimentally studied secondary structure elements.

RNA secondary structures have critical functions not only in RNA viruses, but also in cells. Examples abound and include mRNAs, rRNAs, tRNAs, RNAse P, and tmRNA (10Sa RNA of *Escherichia coli*) (Birchmeier et al., '83; Aziz and Munro, '87; Graves et al., '87; Easterwood and Harvey, '97; Felden et al., '97). However, they represent only one of many forms of suborganismal organization. A number of recent results suggest that the evolution of robustness may be a general phenomenon also on higher levels of biological organization, such as that of genetic networks. (Nowak et al., '97; Stearns and Kawecki, '94; Wagner et al., '97; Wagner, '99). This raises a question about the general preconditions under which mutational robustness might evolve in biological systems. There are at least three such preconditions.

First, the system under consideration must be under strong stabilizing selection. Otherwise, robustness would incur no selective advantage. Second, there must be alternative ways to achieve the same biological function with different degrees of robustness. Possible examples run the gamut of biological organization. They include different amino acid sequences folding into polypeptides of the same function which vary only in their sensitivity to amino acid replacements, as well as alternative interaction patterns among the gene products in a developmental pathway, where each pathway architecture leads to the same patterning process, but has differential susceptibility to mutations. Beyond the observation that polygenic systems with non-additive gene interactions allow such alternative organizations (Wagner, '96; Wagner et al., '97), little is known about this prerequisite for the evolution of robustness.

Third, the amount of variation in robustness generated by mutations has to be sufficiently large for selection to be effective. The reasons lie in the fact that the evolution of robustness is a second order phenomenon: it is not the robust system itself, but its offspring that is better adapted, on average. This requires high mutation rates μ or large (effective) population sizes N_e; for the evolution of robustness via redundant gene functions, for instance, $N_e\mu >> 1$ may be necessary (Wagner, '99). We note that RNA viruses are an ideal study system in this regard, because of their enormously high mutation rates of up to 10^{-3} per position and replication (Drake, '93; Domingo et al., '95).

Evolution of robustness, if widely occurring, has profound implications for our understanding of molecular evolution. From a pragmatic perspective, it provides a good candidate mechanism for an evolving fraction of neutral vs. non-neutral mutations. Everything else being equal, this im-

plies an evolving neutral mutation rate, and might necessitate an even more careful choice of molecules for molecular systematic studies.

Perhaps more profound is the emphasis it puts on the role of the genetic context in which mutations occur. Similar to the effect of a point mutation in an RNA secondary structure element, which depends on the local (robust or non-robust) structural context, the effect of a mutation in a developmental gene will depend on other genes in the developmental pathway (the genetic "background"). While there are clear examples of mutations that will be deleterious in almost any context, the fitness effects of most mutations may be context-dependent. If context is indeed of the essence, then we will have to rethink time-honored and gene-centered notions in population genetics, such as that of a "genetic load" (Crow and Kimura, '70), or the very notion of a "deleterious mutation."

ACKNOWLEDGMENTS

We thank Walter Fontana and Martijn Huynen for invaluable discussions. Computational support by the Albuquerque High Performance Computing Center (AHPCC) is also gratefully acknowledged.

LITERATURE CITED

Aziz N, Munro HN. 1987. Iron regulates ferritin mRNA translation through a segment of its 5′ untranslated region. Proc Natl Acad Sci USA 84:8478–8482.

Baudin F, Marquet R, Isel C, Darlix JL, Ehresmann B, Ehresmann C. 1993. Functional sites in the 5′ region of human immunodeficiency virus type 1 RNA form defined structural domains. J Mol Biol 229:382–397.

Birchmeier C, Folk W, Birnstiel ML. 1983. The terminal RNA stem-loop structure and 80 bp of spacer DNA are required for the formation of 3′ termini of sea urchin H2A mRNA. Cell 35:433–440.

Bonhoeffer S, McCaskill JS, Stadler PF, Schuster P. 1993. RNA multi-structure landscapes: a study based on temperature dependent partition functions. Eur Biophys J 22:13–24.

Bowie JU, Reidhaar-Olson JF, Lim WA, Sauer RT. 1990. Deciphering the message in protein sequences: tolerance to amino acid substitution. Nature 247:1306–1310.

Brinton MA, Fernandez AV, Dispoto JH. 1986. The 3′-nucleotides of flavivirus genomic RNA form a conserved secondary structure. Virology 153:113–121.

Cadigan KM, Grossniklaus U, Gehring WJ. 1994. Functional redundancy: the respective roles of the two sloppy paired genes in *Drosophila* segmentation. Proc Natl Acad Sci USA 91:6324–6328.

Crow JF, Kimura M. 1970. An introduction to population genetics theory. New York: Harper and Row.

Dayton ET, Konings DA, Powell DM, Shapiro BA, Butini L, Maizel JV, Dayton AI. 1992. Extensive sequence-specific information throughout the CAR/RRE, the target sequence of the human immunodeficiency virus type 1 Rev protein. J Virol 66:1139–1151.

Domingo E, Holland JJ, Biebricher CK, Eigen M. 1995. Quasispecies: the concept and the word. Gibbs A, Calisher CH, García-Arenal G, editors. In: Molecular basis of virus evolution. Cambridge: Cambridge University Press. p 171–180.

Drake JW. 1993. Rates of spontaneous mutations among RNA viruses. Proc Natl Acad Sci USA 90:4171–4175.

Dun RB, Fraser AS. 1958. Selection for an invariant character—"vibrissa number"—in the house mouse. Nature 1018–1019.

Easterwood TR, Harvey SC. 1997. Ribonuclease P RNA: models of the 15/16 bulge from *Escherichia coli* and the P15 stem loop of *Bacillus subtilis*. RNA 3:577–585.

Felden B, Himeno H, Muto A, McCutcheon JP, Atkins JF, Gesteland RF. 1997. Probing the structure of *Escherichia coli* 10Sa RNA (tmRNA). RNA 3:89–103.

Fontana W, Konings DAM, Stadler PF, Schuster P. 1993a. Statistics of RNA secondary structures. Biopolymers 33:1389–1404.

Fontana W, Stadler PF, Bornberg-Bauer EG, Griesmacher T, Hofacker IL, Tacker M, Tarazona P, Weinberger ED, Schuster P. 1993b. RNA folding and combinatory landscapes. Phys Rev E 47:2083–2099.

Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH. 1986. Improved free-energy parameters for predictions of RNA duplex stability. Proc Natl Acad Sci USA 83:9373–9377.

Graves RA, Pandey NB, Chodchoy N, Marzluff WF. 1987. Translation is required for regulation of histone mRNA degradation. Cell 48:615–626.

Hartl DL, Dykhuizen DE, Dean AM. 1985. Limits of adaptation: the evolution of selective neutrality. Genetics 11: 655–674.

He L, Kierzek R, Santa Lucia J, Walter AE, Turner DH. 1991. Nearest-neighbor parameters for GU mismatches. Biochemistry 30.

Hofacker IL, Fekete M, Flamm C, Huynen MA, Rauscher S, Stolorz PE, Stadler PF. 1998. Automatic detection of conserved RNA structure elements in complete RNA virus genomes. Nucleic Acids Res 26:3825–3836.

Hofacker IL, Fontana W, Stadler PF, Bonhoeffer S, Tacker M, Schuster P. 1994. Fast folding and comparison of RNA secondary structures. Monatsh Chem 125:167–188.

Hofacker IL, Huynen MA, Stadler PF, Stolorz PE. 1996a. Knowledge discovery in RNA sequence families of HIV using scalable computers. In: Proceedings of the 2nd international conference on knowledge discovery and data mining. Portland, OR: AAAI Press. p 20–25.

Hofacker IL, Huynen MA, Stadler PF, Stolorz PE. 1996b. RNA folding and parallel computers: the minimum free energy structures of complete HIV genomes. Tech. rep. Santa Fe, NM.

Hofacker IL, Stadler PF. 1999. Automatic detection of conserved base pairing patterns in RNA virus genomes. Comput Chem (in press).

Hogeweg P, Hesper B. 1984. Energy directed folding of RNA sequences. Nucleic Acids Res 12:67–74.

Honda M, Brown EA, Lemon SM. 1996. Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. RNA 2:955–968.

Huynen MA, Gutell R, Konings DAM. 1997. Assessing the reliability of RNA folding using statistical mechanics. J Mol Biol 265:1104–1112.

Huynen MA, Konings DAM, Hogeweg P. 1993. Multiple coding and the evolutionary properties of RNA secondary structure. J Theor Biol 16:251–267.

Jackson RJ, Kaminski A. 1995. Internal initiation of translation in eukaryotes: the picornavirus paradigm and beyond. RNA 1:985–1000.

Jaeger JA, Turner DH, Zuker M. 1989. Improved predictions of secondary structures for RNA. Proc Natl Acad Sci USA 86:7706–7710.

Joyner AL, Herrup K, Auerbach BA, Davisb CA, Rossant J. 1991. Subtle cerebellar phenotype in mice homozygous for a targeted deletion of the en-2 homeobox. Science 25:1239–1243.

Kacser H, Burns JA. 1981. The molecular basis of dominance. Genetics 97:639–666.

Lück R, Steger G, Riesner D. 1996. Thermodynamic prediction of conserved secondary structure: application to the RRE element of HIV, the tRNA-like element of CMV, and the mRNA of prion protein. J Mol Biol 258:813–826.

Mandl CW, Holzmann H, Meixner T, Rauscher S, Stadler PF, Allison SL, Heinz FX. 1998. Spontaneous and engineered deletions in the 3′-noncoding region of tick-borne encephalitis virus: construction of highly attenuated mutants of flavivirus. J Virol 72:2132–2140.

Nowak MA, Boerlijst MC, Cooke J, Maynard-Smith J. 1997. Evolution of genetic redundancy. Nature 38:167–171.

Olsthoorn RC, Licis N, vanDuin J. 1994. Leeway and constraints in the forced evolutions of a regulatory RNA helix. EMBO J 13:1994.

Rendel JM. 1979. Canalization and selection. In: Thompson JN Jr, Thoday JM, editors. Quantitative genetic variation. New York: Academic Press. p 139–156.

Schuster P, Fontana W, Stadler PF, Hofacker IL. 1994. From sequences to shapes and back: a case study in RNA secondary structures. Proc R Soc Lond B: Biol Sci 255: 279–284.

Schuster P, Stadler PF. 1998. Sequence redundancy in biopolymers: a study on RNA and protein structures. In: Myerrs G, editor. Viral regulatory structures, vol. XXVIII of Santa Fe Institute studies in the sciences of complexity. Reading, MA: Addison-Wesley. p 163–186.

Schuster P, Stadler PF, Renner A. 1997. RNA structure and folding: from conventional to new issues in structure predictions. Curr Opin Struct Biol 7:229–235.

Sokal RR, Rohlf FJ. 1981. Biometry. New York: Freeman.

Stearns SC, Kawecki TJ. 1994. Fitness sensitivity and the canalization of life history traits. Evolution 48:1438–1450.

Tacker M, Stadler PF, Bornberg-Bauer EG, Hofacker IL, Schuster P. 1996. Algorithm independent properties of RNA secondary structure prediction. Eur Biophys J 25:115–130.

Tautz D. 1992. Redundancies, development and the flow of information. Bioessays 14:263–266.

Thompson JD, Higgs DG, Gibson TJ. 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties, and weight matrix choice. Nucleic Acids Res 22:4673–4680.

Waddington CH. 1957. The strategy of the genes. New York: Macmillan.

Wagner A. 1996. Does evolutionary plasticity evolve? Evolution 50:1008–1023.

Wagner A. 1999. Redundant gene functions and natural selection. J Evol Biol 12:1–16.

Wagner GP, Booth G, Bagheri-Chaichian H. 1997. Towards a population genetic theory of canalization. Evolution 51: 329–347.

Wallner G, Mandl CW, Kunz C, Heinz FX. 1995. The flavivirus 3′-noncoding region: extensive size heterogeneity independent of evolutionary relationships among strains of tick-borne encephalitis virus. Virology 213:169–178.

Weintraub H. 1993. The MyoD family and myogenesis: redundancy, networks, and thresholds. Cell 75:1241–1244.

Wuchty S, Fontana W, Hofacker IL, Schuster P. 1999. Complete suboptimal folding of RNA and the stability of secondary structures. Biopolymers 49:145–165.

Zuker M, Sankoff D. 1984. RNA secondary structures and their prediction. Bull Math Biol 46:591–621.