Transaction Number: 507364

Call #: 

Location: SMITH PER

**Article Information**

*Journal Title:* Biochemistry and Molecular Biology International

*Volume:* 32  *Issue:* 5  *Month/Year:* 1994  *Pages:* 879-887

*Article Author:* 

*Article Title:* The Theory of Genotypic Selection

**Loan Information**

*Loan Title:* 

*Loan Author:* 

*Publisher:* 

*Place:* 

*Date:* 

*Imprint:* 

**Customer Information**

*Username:* piecze

George Pieczenik
Rutgers Faculty - Science/Math

*Article Delivery Method:* Mail to Address

*Loan Delivery Method:* Hold for Pickup

*Electronic Delivery?* Yes

George Pieczenik, Ph.D.

Department of Biochemistry and Microbiology, Lipman Hall 217, Cook College, Rutgers University, New Brunswick, N.J. 08903, U.S.A.

Received January 13, 1994
Received after revision, February 17, 1994

1. Summary

One aspect of the Theory of Genotypic Selection states that G:U base pairs are selected for function and structure. RNA secondary structures stabilised by G:U base pairs are suggested to be involved in the evolution of attenuated strains of polio virus. Asymmetric mutations (C ⇔ U and A ⇔ G) and convergent evolutionary solutions are explained as a direct consequence of G:U base pairing. The possibility of a naturally occurring attenuated strain of HIV-1 is predicted.

2. Introduction

The common understanding is that when most people have looked at molecular evolution they have only, or mainly, considered changes in coding sequences. We believe that the selective process can operate at the level of RNA secondary structure. This direct selection on RNA structures can be considered another example of “genotypic selection” or “genotype selection” (1,2)

A possible example of this process is the generation of certain attenuated strains of poliovirus (3) The region responsible for neurovirulence in polio virus strain 3 and strain 1 have been identified. They are a C to U change at position 472 for strain 3 (4,5) and a A to G change at position 483 for strain 1 (6). Four different RNA structures have been presented for the neurovirulence region of the poliovirus strain 3 and one structure for strain 1. They are summarized in figure 1.

Figure 1 shows the first proposal of a base pair of the RNA secondary structure of the pathogenic Leon strain mutating into the Sabin type 3 attenuated strain in which a C:G base pair (472C:G482) mutates to a U (5). The secondary structure is not preserved in this model of attenuation. Some of the same authors (7) proposed another secondary structure for this neurovirulence region three years later. This time they had the base pair 472C:G627 mutating to the base pair 472U:G627. In this model of attenuation, the secondary structure is maintained with a C:G base pair replaced by a G:U base pair.
In 1989, Skinner, M.A. et al. (8) developed another secondary structure using a similar coding algorithm to the previous models which in addition, incorporated chemical modification and enzymatic cleavage constraints. The structural change correlating to attenuation for strain 3 is a 472C:G537 base pair evolving into a 472U:G537 structure. The structural change correlating to viral attenuation for strain 1 is 483A:U528 base pair to 483U528. In both cases, the structure is maintained by a G:U base pair. Their own CV1 enzymatic cleavage data suggests that in their model the sequence CG at position 464-465 should be in a base paired configuration. Their data refutes the previous secondary structure model (7), which does not accord with the chemical modification of positions 469,470 and 475. In this structure, only the adenine at position 476 is exposed for dimethylsulphate modification.

Figure 1, also, shows a different base pair mutation as a consequence of the secondary structure proposed by Pilipenko et al. (9). They proposed the 472C:G537 base pair to 472U:G537 base pair as the transition responsible for the attenuation of neurovirulence. Their chemical modification data and enzymatic cleavage data is confirms much of their structure. However, a CV1 cleavage shown between 535 C and 536 G is inconsistent with the single looped-out nucleotide proposed. CV1 cleavage specificity
suggests that this nucleotide should be part of a base paired region. Skinner’s (8) CV1 cleavage data is compatible with Piipenko’s (9) base pairing at 465G:C541 and 464C:G5442 but not the base pairing shown at 466G:C540.

Figure 1, ref.(10), also, shows the region of secondary structure and mutation for the Sabin 1 strain evolution from the parent Mahoney strain and into the daughter revertant strains after passage in the human gut. This model was recently proposed by Muzyczynko et al. (10). The model is similar to the previous Piipenko model for strain 1(9). The evidence is more functional and is based on coupled mutation data. The coupled data is the simultaneous concerted changes at position 483 and 528 (numbering according to the type 3 strain, a +3 difference). Since the base pairing between nucleotides 483 and 528 was restored in all cases, the restoration could have been due to either a G483 to A, a true back reversion, or a U528 to C, a second-site mutation. Early postvaccination, “A-U” and “G-C” variants of the virus can be isolated.

3. Previous Methods

All of these previously proposed secondary structure models (11) used variations of an optimum base pairing algorithm (12) weighted with various constraints. This algorithm as used, even with various energy weighting (13) cannot generate the “cloverleaf” model for tRNA (14). This algorithm will generate the maximum base paired structure of a tRNA as a “pin” and not as a “cloverleaf”. However, much as the “cloverleaf” model of tRNA was derived by looking for a common structure among many types and variants of tRNA, some of these previous models used taxonomic comparison data. Rivera et al. (7) used comparisons with enteroviruses and rhinoviruses to establish conserved features. Unfortunately, the conserved features in the region of interest for neurovirulence were not compatible with DM chemical modification data. Skinner et al.(8) also derived a consensus model for polioviruses, coxsackieviruses and rhinoviruses using published sequence data and that they suggest that their “previous minimal energy secondary structure model for poliovirus was not conserved among these other viruses.” This consensus model is not completely compatible with their own CV1 cleavage data. For example, an extended single-stranded region from 444 to 470 is shown as susceptible to CV1 cleavage at 5 separate sites.

4. Results

There are several types of additional constraints that can be combined with these computer algorithms that allows one to generate tRNA clover-leaf structures, a priori, as well as other known structures. The observation exists that optimum secondary structures cannot be generated as the optimum pairing of individual bases. The optimum base-pairing algorithm will generate a “pin” structure for tRNA. However, the optimum combinatorial of potential base pairing regions (regions 4-5 nucleotides long for example) will generate the tRNA cloverleaf, a priori. This constraints allows one to analyse large RNA structures in real time.

Another constraint is the observation that there is a complementary positioning of true palindromic sequences to base pairing sequences. This single stranded mirror image
symmetry (15) is often found as part of the single-stranded regions of a stem-loop RNA structure (15,16). In Figure 2, we propose another RNA base pairing structure that combines constraints previously discussed. This structure has a 5' stem that is a true palindrome and a loop that is a true palindrome. The 5' stem sequence of 467 to 477 i.e. CUAAUCCUAAC is an almost perfect true single-stranded palindrome. The sequence of the loop itself 474-480 i.e. UAACCAU is, also, an almost perfect true single-stranded palindrome.

\[
\begin{array}{c|c|c}
\text{NEUROGENIC} & \text{NON-NEUROGENIC} & \text{NEUROGENIC} \\
\text{LEON/37} & \text{SABIN 3} & \text{REVERTANT} \\
C & C & C \\
A & A & A \\
U & U & U \\
C & G & G \\
\downarrow & \downarrow & \downarrow \\
\text{NON-NEUROGENIC} & \text{SABIN 1} & \text{SABIN 3} \\
C & G:C & G:U \\
A & A & A \\
U & U & U \\
C & C & C \\
\end{array}
\]

\[
\begin{array}{c|c|c}
\text{472} & \text{482} & \text{483} \\
C & U & C \\
A & A & A \\
G & G & G \\
\end{array}
\]

Figure 2. Stem-loop structure involved in attenuation. A true palindromic sequence (italics) is one side of the stem and a terminator sequence as part of another palindromic sequence is the loop of the stem-loop. The same stem-loop structure is proposed to be involved in the attenuation and reversion of Sabin strain 1 and Sabin strain 3. The \(\square\) represents G:C base pairing with 3 hydrogen bonds. The \(\bigboxdot\) represents A:U base pairing with 2 hydrogen bonds. The \(\bigboxdot\) represents G:U base pairing with proposed 1 hydrogen bond.
An additional constraint is the single stranded exposure of the sequence UAA in the loop of this structure. Because UAA, UGA, UAG are the only triplet codons (single stranded RNA) recognized by proteins (the release factors) as opposed to RNA, a constraint on many RNA sequences, that have looped regions and are recognised by proteins, contain one or more of these triplets in the looped out region. This follows from a general conservatism in evolution, and that, given the importance and primordial origin of release factors and their ability to recognise the terminator triplet RNA sequences precisely: then, most RNA protein recognition factors descended from these sequences will recognise these sequences, probably as single-stranded sequence. Therefore, this structure can be viewed as a set of constraints, that are then optimally based paired. The constraints are 5’ true palindrome region - true palindrome and terminator codon loop region - base pairing 3’ region.

In this structure, the 472C:G482 base pair mutates to a 472U:G482 base pair for strain 3 and the 471U:A483 mutates to 471U:G483 base pair for strain 1. In both cases, the secondary structure is conserved by the attenuating mutation. The secondary structure also fits the chemical modification data as well as the enzymatic cleavage data presented in the previous models. This model and River et al.’s (7) model have one base pair, Pilipenko et al.’s(9) model has three base pairs. There are two CV1 cleavage sites. Their DM data shows modifications at the adenines at 475,476 and at the adenines at 469,470 but not at adenine 479. Skinner et al. (8) and Pilipenko et al. (9) show adenine 479 base paired to U532.

Figure 1 summarises the RNA secondary structure models of polio virus attenuation for strain 3 and strain 1. All the models, except Evans et al. (5), show a G:U base pair in the secondary structure as part of the viral attenuation model.

Therefore, the preservation of RNA secondary structure by G:U base pairing is part of several models for the evolution of the attenuated strains of Sabin polio virus strain 3 and Sabin polio virus strain 1. This region of the polio virus is untranslated.

It should be noted, that in recent data presented by Muzychenco et al. (10) on the evolution of strain 2 of the Sabin vaccine i.e. strain P712ch2ab; the evolutionary change for virulence is a change of A to G with a 398U:A481 base pair (or 401U:A484 using the type 3 numbering) evolving to a 398U:G481 base pair. This makes the G:U base pair containing variant of strain 2 the virulent variant. This is opposed to the avirulent Sabin strains 3 and strain 1 containing the G:U base pair. However, Minor and Dunn, (17) have previously shown that this same change occurs in healthy vaccines.

Muzychenco et al. (10) point out that most isolates of changes to virulence are coupled changes of U to C and A to G i.e. U:A to C:G, at 398C:G481. Only two and one-half cases (one had both U and C) out of fourteen cases of virulence observed with an uncoupled transition of A to G, without the respective U to C change. This observation
allows a G:U base pair and a C:G base pair to be correlated with virulence. Evans et al. (5) assayed the neurovirulence of various strains of type 3 revertants. One such strain, DM2 contains a U at position 472 that is characteristic of the non-neurogenic Sabin strain. When DM2 was assayed in a standard WHO neurovirulence test, it had a higher lesion score than the Sabin vaccine strain. Evans et al. interpreted this to mean that "DM2... contained a small proportion of virus with C at position 472 which was undetectable by the sequencing methods used." The virulence observed for the strain 2 "G:U" variants could be a contamination of 398C:G481 (or 401C:G484 in type 3 numbering).

Is there any evidence for the functional, as opposed to just the structural existence of G:U base pairing interactions?

One evolutionary consequence of G:U base pairing interactions is possible tRNA anti-codon flipping (18) and a ribosome free protein synthesising system (19). Clark, B.F.C. (20) proposed that the RNA double-helix can internally accommodate a single G:U base pair without serious distortion.

We propose, for the first time, that another direct functional consequence of having RNA-RNA interactions with G:U base pairing is a transition mutational asymmetry of C to U and A to G.

That is, one can give an arrow to evolution and give a preferred mutation direction i.e. that is C $\rightarrow$ U and A $\rightarrow$ G.

Figure 3 shows that given a hypothetical structure with equally frequent G:C, A:U and G:U base pairs, and allowing for equally frequent mutations that preserve the RNA structure; one generates an asymmetry of mutation direction. C to U and A to G transitions are preferred 2 to 1 over U to C and G to A transitions.

Brian Clark (20) showed that the frequencies of G:C and A:U base pairs are higher than G:U in naturally occurring RNA structures and this would suggest an even greater asymmetry of mutation.

Robertson, H. and Jeppesen, P. (21) first noted an asymmetry of mutation direction in the three related RNA bacteriophage, R17, f2 and MS2. They speculated that "If changes from U to C or C to U were to reflect some natural evolutionary characteristic of spontaneous mutations, then this asymmetry might allow us to define a historical relationship between the three phages. For example, if U to C changes predominate, we might conclude that MS2 and f2 diverged along a pathway on which R17 arose earlier. However, if C to U changes are characteristic, the asymmetry in our data could only suggest some sort of convergence of f2 and MS2 along separate pathways, with R17 arising when this process was nearly complete."
5. Discussion

If our model of the preservation of RNA structures by G:U base pairing is correct and the extensive secondary structure in RNA bacteriophage is correct; then, Robertson and Jeppesons (21) second proposition, that of the convergence of f2 to MS2 to R17 is, therefore, the most likely to be correct. As convergent evolutionary solutions are much rarer than the classical divergent (descent from a common ancestor) solutions, this is a very important piece of evidence for selection at the RNA level for G:U base pairing stabilising RNA structures and, therefore, genotypic or genotype selection.

More recently, Goodenow, M. et al., (22) have shown that the most common base change observed (in rapidly evolving HIV-1 isolates) was indeed the G → A transition. They go on to suggest that either a rG:dT mismatch or a rA:dC mismatch will create this bias and that “G:T and G:U mismatches, respectively, in DNA and RNA, are known to be less
destabilizing that A:C pairs.” They also suggest that deamination of C to U as a possible mechanism.

Kestler, H.W. et al. (23) has shown that a single nucleotide change in the nef gene of a cloned form of SIVmac from a terminator (UAA) to codon such as CAA (GLN) can result in maintaining a high viral load for “full pathologic potential” in infected rhesus, while deleting the nef gene had no “detectable effect on virus replication in cultured cells”. In this case, as opposed to the poliovirus, (in which the C to U change is in an untranslated region) this C to U transition allows for the translation of the nef protein, which can be correlated to clinical pathology, but not to cytopathology.

If the preservation of RNA structures by G:U base pairing can be correlated to the evolution, convergent or otherwise, of attenuated strains of RNA virus, then one can calculate the maximum “target-size” for “attenuation”. Since, it is known that a half a turn of an RNA or DNA double helix is stable by itself, two half turns surrounding a G:T or G:U base pair would be stable. Therefore, the maximum base pairing or interacting RNA structure involved in attenuation would be in the magnitude of 18-22 nucleotides long.

Given that the poliovirus is 7,431 nucleotides long, a possible mutation frequency for an attenuated strain of polio would be 22/7431 or 2.7 x 10^-3. As HIV is 9213 nucleotides long, a possible mutation frequency for an attenuated strain would be 22/9213 or 2.2 x 10^-3.

Such an attenuated strain of HIV-1 may have been identified at this frequency, intentionally, (24,25), unintentionally (26) at an unknown frequency, and during a retrospective search of a transfusion from a single donor into six recipients (27).

This paper is dedicated to the memory of my mother, Teodora Janowska Pieczenik and my father, Dr. Srul David Pieczenik.

References

change in the nef gene which as CAA (GLN) can be correlated to the untranslating region. In this way which can be correlated to the untranslating region. This was observed for the case of the virus, then one can proceed to look for a codon or a half turn which is binding the surrounding G:T or G:G base pairs or interacting RNA nucleotides. In most cases, the nucleotides long.