The Nucleotide Sequence at the 5'-Terminus of the Qβ RNA Minus Strand

H. M. Goodman*, M. A. Billeter, J. Hindley, and C. Weissmann

INSTITUT DE BIOLOGIE MOLECULAIRE, UNIVERSITÉ DE GENÈVE, INSTITUT FÜR MOLEKULARBIOLINIQUE DER UNIVERSITÄT ZURICH AND DEPARTMENT OF BIOCHEMISTRY, UNIVERSITY OF BRISTOL

Communicated by V. Prelog, July 23, 1970

Abstract. The sequence of the first 52 nucleotides at the 5'-end of the Qβ minus strand has been determined and found to be complementary in an anti-parallel fashion to the 3'-terminal region of the Qβ plus strand. There are few similarities between the corresponding sequences of Qβ plus and minus strands; however, both have a hydrogen-bonded loop close to the 5'-end. The sequence of the 3'-terminal region of the plus strand was deduced: there are no termination signals within the last 32 nucleotides. A notable homology exists between the 3'-ends of Qβ and R17 RNA.

Considerable progress has recently been made in elucidating the nucleotide sequences of several biologically important regions of the RNA from the bacteriophages R171-4 and Qβ.7 We have described previously a new method which provides a systematic approach to the determination of the structure of Qβ RNA. In this method, purified Qβ-specific RNA polymerase ("Qβ replicase")8 is used to synthesize radioactive segments of RNA in synchronized short-time reactions using [α-32P]-labeled nucleoside triphosphates as substrates. By an appropriate choice of incubation conditions it is possible to control the length of the segments, and to obtain essentially any desired labeling pattern by the introduction of either one or any combination of several radioactive nucleoside triphosphates at various times during synthesis. The nearest neighbor data and the information obtained by sequential labeling of these segments greatly facilitates the subsequent sequence analysis.

This approach was already used to determine the nucleotide sequence of the first 175 nucleotides from the 5'-terminus of the Qβ RNA plus strand synthesized in vitro.5 We report here the application of this method to the elucidation of the sequence of a segment 52 nucleotides long from the 5'-end of the minus strand of Qβ RNA; this has allowed us to deduce the 3'-terminal sequence of the plus strand.

Materials and Methods. Synthesis in vitro and analysis of the minus strand segment. Qβ replicase4 uses Qβ viral RNA (plus strands) as a template for the synthesis of a complementary minus strand.9 Synthesis starts with the 5'-terminal nucleotide and proceeds toward the 3'-end of the RNA.10 The experimental conditions used for the synthesis of minus strands in vitro are identical to those described previously for plus strand synthesis except that, in this case, plus strands were used as template and the host cell
factors\textsuperscript{12,18} required for minus strand synthesis were added. From earlier studies\textsuperscript{14,16} on the 3'-terminus of the plus strand, we expected the first fourteen nucleotides at the 5' end of the Q8 RNA minus strand to be purines; synchronization of synthesis was therefore performed by incubating enzyme, template, and factors with GTP and ATP. As in earlier experiments\textsuperscript{5,6,7} this permits initiation and synthesis of the chain up to the point where the first pyrimidine nucleotide (a C at position 15) is required. The addition of CTP and UTP after this incubation period starts the synchronous chain elongation.

One-half of the 3S product obtained after a 10 sec incubation at 20°C was digested with T\textsubscript{1} ribonuclease and the other half with pancreatic ribonuclease. The resulting oligonucleotides were separated by two-dimensional electrophoresis (Fig. 1).\textsuperscript{16} 

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig1.png}
\caption{Autoradiograph of (a) T\textsubscript{1} and (b) pancreatic RNase digestion products of the 10-sec minus strand segment labeled with 4 nucleoside triphosphates. Synthesis and analysis of the product was performed as outlined in the text and as described.\textsuperscript{17} (a) Products of T\textsubscript{1} RNase digestion. For identification of the spots corresponding to oligonucleotides in positions 1-32, see Table 1. In addition, 10 = AG; 18 = CCAAUUG; 20 = AU (AAUU)\textsubscript{2} (AUC) (CAG); 22 = UAAUUG! (Spot 26 (pppGp) ran off the paper.)

(b) Products of pancreatic RNase digestion. Spots not identified in Table 2 are 1 = U; 2 = U; 3 = GC; 6 = GGC; 8 = AGGC; 9 = GAU; 23 = AU. (Spots 30 (C) and 32 (AC) ran off the paper.)}
\end{figure}

The nucleotide sequences of these end-products (see Tables 1 and 2) were determined from nearest neighbor and total composition data obtained from segments labeled with each of the four nucleoside triphosphates separately and with all four simultaneously. The details of this type of analysis have already been described.\textsuperscript{7,17} Additional experiments were necessary to complete the sequences in only two cases, namely: U\textsubscript{2} ribonuclease\textsuperscript{18} and snake venom phosphodiesterase\textsuperscript{19} digestion of T17; and snake venom phosphodiesterase digestion, as well as sequential oxidation and elimination,\textsuperscript{14} in the case of P20.
TABLE 1.  \(T_1\) ribonuclease products of the 10-sec product (minus strand).

<table>
<thead>
<tr>
<th>Spot</th>
<th>Moles</th>
<th>Standard time of appearance, sec</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>2</td>
<td>&lt;5</td>
<td>G[G]</td>
</tr>
<tr>
<td>1c</td>
<td>1</td>
<td></td>
<td>G[C]</td>
</tr>
<tr>
<td>1d</td>
<td>2</td>
<td></td>
<td>G[A]</td>
</tr>
<tr>
<td>4a</td>
<td>2</td>
<td>&lt;5</td>
<td>AG[G]</td>
</tr>
<tr>
<td>4b</td>
<td>1</td>
<td>?</td>
<td>AG[U]</td>
</tr>
<tr>
<td>4c</td>
<td>2</td>
<td>&lt;5</td>
<td>AG[A]</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>&lt;5</td>
<td>CAG[A]</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>&lt;5</td>
<td>CAAAG[C]</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>8</td>
<td>UAAG[A]</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>8</td>
<td>AUCCCCUCUCACUCUG[U]</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>11</td>
<td>UAAUUG[U]</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>&lt;5</td>
<td>pppG[G]</td>
</tr>
</tbody>
</table>

Synthesis and analysis of labeled products were carried out as outlined in the text. The terminal nucleotide in square brackets is deduced from nearest neighbor data and is not part of the oligonucleotide.

Table 2.  Pancreatic ribonuclease products of the 10-sec product (minus strand).

<table>
<thead>
<tr>
<th>Spot</th>
<th>Moles</th>
<th>Standard time of appearance, sec*</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>1</td>
<td>10</td>
<td>U[G]</td>
</tr>
<tr>
<td>1c</td>
<td>3</td>
<td></td>
<td>U[C]</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>&lt;5</td>
<td>AAAGC[A]</td>
</tr>
<tr>
<td>7a</td>
<td>1</td>
<td>11</td>
<td>GU[G]</td>
</tr>
<tr>
<td>7b</td>
<td>1</td>
<td></td>
<td>GU[A]</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>&lt;5</td>
<td>AGAU[C]</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>9</td>
<td>AAGAGU[A]</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>5  pppGGGAGGAGAGGGG[A]</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td></td>
<td>AAGU[U]</td>
</tr>
<tr>
<td>30a</td>
<td>1</td>
<td></td>
<td>C[G]</td>
</tr>
<tr>
<td>30b</td>
<td>2</td>
<td></td>
<td>C[U]</td>
</tr>
<tr>
<td>30c</td>
<td>3</td>
<td></td>
<td>C[C]</td>
</tr>
<tr>
<td>30d</td>
<td>1</td>
<td></td>
<td>C[A]</td>
</tr>
<tr>
<td>32b</td>
<td>1</td>
<td></td>
<td>AC[U]</td>
</tr>
</tbody>
</table>

*Because the time of appearance was studied only on products labeled with \([^{32}P]\)GTP, this parameter is known only for the oligonucleotides labeled by GTP.

Tables 1 and 2 also list a "time of appearance" for all oligonucleotides labeled by GTP. This parameter was defined previously as the time at which half a mole of the oligonucleotide is present, as interpolated from a sequential labeling experiment.\(^7\)\(^17\)

Results and Discussion.  Primary structure.  A unique sequence for the 3S product can be deduced from direct sequence overlaps between the pancreatic and \(T_1\) oligonucleotides, nearest neighbor data, and the time of appearance of label (Fig. 2). Comparison with the sequence of the first 52 nucleotides at the 5'-end of the viral (plus) strand determined previously\(^8\) shows that the first three nucleotides, pppGGG... , are the same in both cases (Fig. 2). It is not certain whether this, as well as a further similarity (segment 38–50 of the plus strand and 34–46 on the minus strand with a sequence ACUNNNNGAGUA in common) is significant. It is also evident that the 5'-end of the minus strand is not the complement of the 5'-end of the plus strand, showing that plus and minus strands are not complementary in a parallel configuration.
Two different sequences have been proposed for the 3'-terminal T₁ oligonucleotide of Q₈ RNA (plus strand), GCCCUCUCUCUCCCA-OH²⁴ and GCCCU-
CUCUCUCCCAA-OH (ref. 15 and personal communication). Comparison of
these two sequences with the first fifteen nucleotides of the 5'-terminal sequence
of the minus strand (Fig. 2) shows that, disregarding the 3'-terminal A, the

5' TERMINAL REGION OF Q₈ MINUS STRAND

[Diagram showing nucleotide sequence]

5' TERMINAL REGION OF Q₈ PLUS STRAND

[Diagram showing nucleotide sequence]

3' TERMINAL SEQUENCE DEDUCED FOR Q₈ PLUS STRAND

...CACAUACUCUACGAGUGAGAGGGAUCUCUUUGCCUCUCUCUCC(A)

Fig. 2. Nucleotide sequence at the 5'-terminus of Q₈ minus strand synthesized in vitro. The oligonucleotides designated with T and P resulted from T₁ and pancreatic RNase digestion, respectively. The numbers written immediately above or below the oligonucleotide designation give the standard time of appearance. The sequence is compared with the corresponding region of the Q₈ plus strand.² The 3'-terminus of the plus strand has been deduced from the 5'-terminus of the minus strand; the presence of a terminal A has been determined by direct analysis.²¹,²⁴

sequence proposed by Gilham²⁴ is the antiparallel complement of the 5'-end of the minus strand. The antiparallel complementarity demonstrated here for the Q₈ plus and minus strand is in agreement with x-ray diffraction data on double-stranded MS2 RNA.²⁰ It has been suggested²¹⁻²³ that synthesis of a minus strand on a plus strand template begins not at the 3'-terminal adenosine but at the penultimate nucleotide (C). This is substantiated by our findings, which show that the minus strand initiates with pppG (rather than pppU), followed by a series of nucleotides complementary to the 3'-terminal plus strand sequence extending from the penultimate C. The question as to the origin of the 3'-terminal A of the plus strand, which can obviously not be specified by the template in a conventional fashion, is discussed elsewhere.²²,²⁴

Given the antiparallel complementarity of plus and minus strands we can deduce the sequence of the 3'-terminal region of the plus strand from the 5'-end of the minus strand, as shown in Fig. 2. None of the three codons known to be signals for polypeptide chain termination in Escherichia coli, UAA (ochre), UAG
(amber), or UGA are found between the 3'-terminus and the 32nd nucleotide from the 3'-end. This complements our previous finding that at least 61 nucleotides from the 5'-end are not part of a viral cistron.

**Secondary structure at the ends of Qβ plus and minus strands.** A remarkable feature of the structure of the plus strand is the occurrence of a loop with a strongly hydrogen-bonded stem, very close to the 5'-end. The sequence data presented in this report suggest that an analogous structure may occur at the 5' terminus of the minus strand. As shown in Fig. 3, the loop of the plus strand

**POSSIBLE SECONDARY STRUCTURE OF Qβ MINUS STRAND**

![Possible Secondary Structure of Qβ Minus Strand]

**POSSIBLE SECONDARY STRUCTURE OF Qβ PLUS STRAND**

![Possible Secondary Structure of Qβ Plus Strand]

**Fig. 3.** Possible secondary structure at the 5'- and 3'-termini of Qβ plus and minus strands. The secondary structure at the 5'-terminus of the plus strand is supported by experimental evidence. Two structures can be written for the 3'-terminus of the plus strand, one of which requires a G-U base pair. The 3'-terminal A of the minus strand was determined directly.

comprises 5, and that of the minus strand 12, nucleotides; both stems consist of 7 nucleotide pairs, predominantly GC. The sequences of the two structures are different. It is expected that the 3'-terminal regions of the plus and minus strand also have well defined secondary structures (Fig. 3).

**Recognition of Qβ specific RNA by Qβ replicase.** Only a few RNA species are known to be accepted as templates by Qβ replicase. Qβ RNA (plus strands), Qβ minus strands, so-called “variants” of Qβ RNA, and a “6S RNA” described recently are replicated by this enzyme; poly C and C-containing copolymers elicit synthetic activity; however, the reaction appears to be limited to one round of complementary synthesis. Neither other viral RNAs, in particular the RNA of phage MS2—a virus related to Qβ, albeit not too
closely—nor ribosomal and tRNA are replicated by \( \Phi Q \) replicase. As far as the evidence goes, there is only one enzyme species in \( \Phi Q \) replicase preparations, and the question arises as to how one enzyme recognizes two different RNAs while discriminating against a host of others. One possibility envisaged earlier\(^9\) was that plus and minus strands had identical sequences at their termini and that these served as recognition sites. A comparison of the terminal regions of \( \Phi Q \) plus and minus strands (Fig. 2) shows that the correspondence between sequences extends to only 3 nucleotides at the 5'-(pppGGG . . .) and 4 at the 3'-terminus (CCCA). While the 3'-terminal A may be removed without affecting the template activity of the plus strand (the minus strand has not been tested), the penultimate C appears to be required for template activity.\(^22\) On the other hand, since different viral RNAs, such as f2,\(^14\) MS2,\(^33\) R17,\(^2\) and TMV,\(^34\) also terminating with CCCA at the 3'-end, are inactive as templates\(^9,36\) it may be concluded that while a sequence...CCC near the terminus may be an absolute requirement for template activity, it cannot be the feature on which the specific relationship between \( \Phi Q \) replicase and its template is based. By the same token it may be argued that positions -23 to -18 (Fig. 4) in the 3'-terminal

\[
\text{...AGCUGCUUGCCUCG, UACCUCUCUCUCUCCCAON R17}\]

\[
\text{...ACUGCUUUACCUCUCUCCCAON Q}\Phi
\]

**Fig. 4.** Comparison of the 3'-terminal regions of \( \Phi Q \) and R17 RNA. The sequences have been aligned from the 3'-end; to obtain maximal homology, one deletion has been allowed for in the R17 sequence reported by Cory et al.\(^46\)

region of \( \Phi Q \) RNA do not constitute the critical recognition site, since this sequence also occurs in the corresponding region of R17 RNA.

Since at least 61 and 32 nucleotides, respectively, at the termini of plus and minus strands do not lie within cistrons, it is possible that these, or even more internal sequences are involved in the recognition function, either by providing some characteristic secondary or tertiary structures (such as the terminal loops described), or by virtue of specific sequences (e.g., regions 38–50 of the plus, and 34–46 of the minus strand). While it is possible that plus and minus strands may possess similar structural features despite their largely different nucleotide sequences, it appears likely that \( \Phi Q \) replicase can in fact recognize the two strands independently. This is indicated by the finding that purified \( \Phi Q \) replicase by itself is capable of using \( \Phi Q \) minus strands as template; however, for \( \Phi Q \) plus strands at least one host-specified protein (found both in uninfected and infected \textit{E. coli}) must be present during initiation.\(^37,38\) \( \Phi Q \) replicase may possess two recognition functions, one of which is dependent on an additional determinant, the host-specified factor. The fact that, in the absence of factor, \( \Phi Q \) replicase can recognize the minus but not the plus strand as template is an additional argument toward the conclusion that the short terminal sequence (CCCA)
common to plus and minus strands (and other viral RNAs) cannot be the critical recognition signal.

**Homology between Qβ and R17.** RNA phages have been divided into three groups on the basis of their serological properties. Of the phages studied in detail, f2, R17, MS2, and M12 belong to group 1, while Qβ is assigned to group 3. Phages belonging to these two groups are not only immunologically unrelated, but also differ in their sizes, densities, and isoelectric points. Moreover, the viral RNAs have different sedimentation coefficients, base compositions, and annealing properties; also, their replicases differ in template specificity and stability. Despite these differences, the coat proteins of Qβ and R17 show similarities in their amino acid sequences. Konigsberg et al. have recently reported that, allowing for some deletions and insertions, the polypeptides of Qβ and f2 (which differs in only one position from R17) can be aligned so that 30 residues occupy identical positions; 62 further residues are related by a one-base change. Moreover, in comparing 74 nucleotides of the 5'-terminal sequence of R17 RNA with 175 residues from the corresponding segment of Qβ RNA, Adams and Cory have noted two regions of homology: (1) The first 22 nucleotides of both RNAs form similar 5'-terminal loops, with identical nucleotides in 15 out of 22 positions; (2) between positions 44–71 in R17 and 101–128 in Qβ, 20 out of 28 nucleotides correspond.

We now find that there is also a notable homology between the 3'-terminal region of Qβ RNA and that recently reported for R17 RNA. If the RNAs are aligned from their 3'-termini, and one deletion is allowed for, 16 out of 24 positions correspond, as shown in Fig. 4. In particular, a sequence of six nucleotides (CUGCUU) is situated at the apex of a loop in both cases.

The homologies described above are compatible with the idea that phages of groups 1 and 3 are derived from a common ancestor. Nevertheless, one cannot as yet exclude the possibility of a convergent evolution governed on the one hand by the relationship of the phage coat proteins to the bacterial surface, and on the other hand by the relationship of the viral RNA to the phage proteins. Further restrictions may be imposed on the RNA structure by the intracellular environment of the host, as well as by demands arising from its function as a template for protein and RNA synthesis.

We thank Mr. Werner Römer, Mrs. Françoise Scholler, and Mrs. Libuse Gross for their valuable technical assistance. This project is supported by grants from the Schweizerische Nationalfonds (to C. W. and to Professor A. Tissières); the British Science Research Council (to J. H.) and NATO (to J. H.); the Jane Coffin Childs Fund (to C. W.); the American Cancer Society Inc. (to H. M. G.) and EMBO (to C. W.).

* Present address: Department of Biochemistry and Biophysics, University of California, San Francisco Medical Center, San Francisco, Calif. 94122.

44 Weber, K., Biochemistry, 6, 3144 (1967).