Nucleic Acids Research, 1994, Vol. 22, No. 3 419–426

Context effects on N6-adenosine methylation sites in prolactin mRNA

Prema Narayan+, Rachael L.Ludwiczak, Edward C.Goodwin and Fritz M.Rottman*
Department of Molecular Biology and Microbiology, Case Western Reserve University, Cleveland, OH 44106, USA

Received September 17, 1993; Revised and Accepted December 28, 1993

ABSTRACT

The methylation of internal adenosine residues in mRNA only occurs within GAC or AAC sequences. Although both of these sequence motifs are utilized, a general preference has been noted for the extended sequence RGACU. Not all RGACU sequences in an mRNA are methylated and the mechanisms that govern the selection of methylation sites in mRNA remain unclear. To address this problem we have examined the methylation of transcripts containing sequences of a natural mRNA, namely, bovine prolactin mRNA. In this mRNA, a specific AGACU sequence in the 3′ untranslated region is the predominant site of methylation both *in vivo* and *in vitro*. The degree to which N6-adenosine methyltransferase recognizes the sequence context of the consensus methylation site was explored by mutational analysis of the nucleotides adjacent to the core sequence as well as the extended regions in which the core element was found. Our results indicate that efficient methylation depends on the extended five nucleotide consensus sequence but is strongly influenced by the context in which the consensus sequence occurs within the overall mRNA molecule. Furthermore, consensus methylation sites present in an RNA duplex are not recognized by the methyltransferase.

INTRODUCTION

Internal N6-methyladenosine (m6A) is an ubiquitous post-transcriptional mRNA modification of all higher eukaryotes and those viruses that replicate in the nucleus (1, 2). Though the functional role of this modification is unclear, studies with methylation inhibitors suggest that m6A is involved in the processing of nuclear RNA (3–7).

Studies of cellular mRNAs and viral RNAs have indicated that N6-adenosine methyltransferase exhibits remarkable sequence specificity *in vivo*. Methylation of internal adenosine residues occurs only at GAC and AAC sequences (the adenosine in **bold** type can be potentially methylated) with a preference for GAC methylation (2). Although GAC or AAC sequences are expected to appear, on average, once every 32 nucleotides in a random RNA sequence, only one to two m6A residues are found per 1000 nucleotides in mRNA. Therefore, only a subset of GAC or AAC motifs are suitable methylation substrates. Examination of individual mRNAs indicates that the distribution of m6A is non-random; methylated adenosines are often clustered in one region of the RNA and are independent of the number of AAC and GAC sequences present and the size of the mRNA (8–12). Early studies by Schibler *et al.* on heterogeneous nuclear RNA indicated that methylation occured in an extended consensus sequence of $N_1(G/A)ACN_2$ with N_1 being a purine 90% of the time and N_2 rarely being a G (13). All of these observations suggested that a sequence longer than 3 nucleotides was required and/or other structural features of RNA were important in imparting specificity.

The precise localization of m6A has been determined *in vivo* for only two mRNAs —the Rous sarcoma virus (RSV) genomic RNA and bovine prolactin (bPRL) mRNA. In RSV genomic RNA, the extended consensus GGACU is most frequently methylated followed by UGACU and AGACU. AAC sequences, although present in RSV genomic RNA, were not found to be methylated (14, 15). In bPRL mRNA, *in vivo* and *in vitro* methylation occurs almost exclusively at a specific AGACU sequence in the 3′ untranslated region despite the presence of 27 GAC and AAC sequences within this mRNA (16). In bPRL precursor mRNA (containing only the last intron) the AGACU sequence within this intron is predominantly methylated both *in vivo* and *in vitro* (7). However, unlike the RSV RNA, AAC sequences are also methylated in intron D of bPRL precursor mRNA. Thus it is likely that an extended AAC consensus sequence is also required for recognition by the methyltransferase.

Mutational analysis of the RSV genomic RNA indicated that methylation is inhibited when there are changes in the GAC core sequence itself or at the +2U position of the AGACU extended consensus sequence, where the methylated A is designated as 0 (15–17). Synthetic RNA substrates of 20 nucleotides in length containing one or more GGACU sequences (and mutations within) have been utilized to test the sequence specificity of methylation *in vitro* (18). These results are similar to the *in vivo* results obtained with the RSV RNA; GGACU was the best

*To whom correspondence should be addressed

+Present address: Department of Biochemistry, Life Sciences Building, University of Georgia, Athens, GA 30602, USA

substrate and AAC sequences were methylated at significantly lower levels. Both these studies confirmed the nature of the extended consensus sequence for m^6A methylation at GAC sites.

Although a consensus sequence of RGACU is an important component in the recognition of target sites for m^6A methyltransferase, it is insufficient to define the methylation sites, especially in the context of a natural mRNA sequence. All RGACU consensus sequences are not methylated equally in mRNA and some sequences are not methylated at all. This implies that additional sequence or structural features beyond this 5-nucleotide consensus sequence (context effects) are important in defining the methylation sites of mRNA. The experiments of Rana and Tuck (19) demonstrate the difficulty in defining an active methylation site. A synthetic RNA transcript was constructed containing 10 copies of a 30 base segment from the RSV genome, such that each segment contained two perfect methylation consensus sequences that exhibited the highest levels of methylation *in vivo*. Though this substrate RNA contained 20 perfect consensus sites in tandem it was a poor substrate for methylation when tested *in vitro*. These results suggest that consensus sequences removed from their normal mRNA environment can be compromised as methylation substrates. Context effects may be a general feature in the recognition of specific signals within mRNA. Indeed, such context effects have been shown to play a role in other RNA processing events, such as in the recognition of consensus splice donor and acceptor sequences in pre-mRNA splicing systems (20–24).

In this report, we have used an *in vitro* methylation system and mutational analysis to examine the primary sequence requirements for AAC methylation, the role of mRNA context in the methylation of naturally occuring consensus sequences in bPRL mRNA and the effect of RNA duplex structure on m^6A methylation. For these studies we have created new methylation sites both at the 3′ end of bPRL mRNA (which contains the major methylation site) and at the 5′ end of this mRNA (which is normally not methylated to detectable levels *in vivo*). Our results indicate that the core AAC sequence is also part of an extended consensus sequence AAACU, that is recognized by the methyltransferase and that efficient methylation of both AGACU and AAACU sequences depends not only on the primary consensus sequence but also on the overall context in which this sequence is located in the mRNA. Furthermore, the ability of the methyltransferase to recognize a consensus sequence is impaired when this sequence is present in a stable RNA:RNA duplex.

MATERIALS AND METHODS

Plasmids

Clones containing only the 3′ terminus of bPRL cDNA were generated by subcloning sequences from the full length bPRL cDNA into the vector pBSM13$^+$ (Stratagene, Inc.). Two oligonucleotides were used to generate a template for the 60 nt fragment of bPRL mRNA by the polymerase chain reaction (PCR). The first is a message-sense oligonucleotide containing the T7 RNA polymerase promoter and 15 nucleotides of bPRL sequence beginning 60 nucleotides upstream of the major polyadenylation site and the second an antisense oligonucleotide containing the Hind III recognition sequence, 50 nucleotides of dT residues and 15 nucleotides complementary to the 3′ end of the bPRL cDNA. The product is a double stranded bPRL template containing the T7 RNA polymerase promoter at the 5′ end, the last 60 nucleotides of the bPRL mRNA, a stretch of 50 dA residues and a Hind III site at the 3′ end. This DNA was cloned into the Sma I and Hind III sites of pEMBL 19 and named bPRL 60. The clone containing the T7 RNA polymerase promoter, 192 nucleotides of the 5′ end of the prolactin cDNA sequence and 50 dA residues was constructed in a similar fashion as the 3′ terminal clone and named bPRL 192.

Mutagenesis

Site specific mutations in the 3′ end of bPRL mRNA were generated by oligonucleotide-directed mutagenesis. The bPRL cDNA containing the T7 RNA polymerase promoter from plasmid bPRL 60 was subcloned into the vector pGP12 and oligonucleotide-directed mutagenesis was performed as described (25). The sequence of the mutant DNAs was confirmed by dideoxy sequencing of plasmid DNA (26). The mutant DNAs were removed from pGP12 and were inserted into pEMBL19 because pGP12 also contains a T7 RNA polymerase promoter.

Site specific mutations in the 5′ end of bPRL RNA (bPRL 192) were generated by PCR using mutant oligonucleotides and the mutant DNA sequences were ligated into the SmaI site of pEMBL19. The sequence of the mutant DNAs was confirmed by dideoxy sequencing (26).

RNA synthesis

Plasmids were linearized and prolactin transcripts were transcribed by T7 RNA polymerase in the presence or absence of G1 cap (m^7GpppG_m) as primer. Radioactively labeled RNA was synthesized in the presence of either low levels of α-[^{32}P]-GTP or α-[^{32}P]-ATP. RNA transcripts were purified by oligo dT chromatography or by gel electrophoresis (27). Uncapped transcripts were treated with calf intestinal alkaline phosphatase (0.02 U/pmole 5′ end) in a buffer containing 50 mM Tris HCl pH 8.5, 0.1 mM EDTA at 37°C for 20 minutes. Removal of the 5′ terminal phosphates renders the RNA incapable of being capped by the nuclear extract. The RNA was subsequently treated with 25 μg proteinase K in a buffer containing 50 mM Tris HCl pH 8.5, 10 mM EDTA, 10 mM NaCl and 0.2% SDS. After a 10 minute incubation at 37°C, the RNA was extracted twice with an equal volume of phenol:chloroform (1:1) and ethanol precipitated.

In vitro methylation

In vitro methylation was performed using a HeLa cell nuclear extract under conditions that had been optimized to give accurate methylation of bPRL mRNA (16). The reactions contained 24 mM Hepes pH 7.9, 20 mM KCl, 2 mM $MgCl_2$, 1.1 mM dithiothreitol, 4% glycerol, 0.04 mM EDTA, 3% polyvinyl alcohol, 2 μM S-[methyl-^{3}H] adenosyl-L-methionine (specific activity 70–80Ci/mmole), 10 μl HeLa cell nuclear extract and 13 pmoles or 26 pmoles of RNA in a final volume of 50 μl. After incubation at 30°C for 10 minutes or 30 minutes, the reaction was stopped by incubation at 65°C for 5 minutes.

Under these conditions both S-[methyl-^{3}H] adenosyl-L-methonine (^{3}H-SAM) and the nuclear extract are in excess. Comparison of the methylation levels of the different substrates was facilitated by the development of a rapid, sensitive and reproducible assay. In this assay the methylated substrate RNA with its poly A tail was purified from the reaction mix by incubating the reactions with oligo dT cellulose. An equal volume of 2× oligo dT loading buffer (20mM Tris HCl pH 7.6, 1.0 M LiCl, 2 mM EDTA, 1.0 % SDS) was added to each reaction

followed by 50 μl of a 10 mg/ml slurry of oligo dT cellulose in oligo dT loading buffer. The tubes were incubated with constant mixing at room temperature for 30 minutes. The contents of each tube were transferred to individual wells of a 96 well millititer filtration plate (Millipore Corporation). The millititer filtration plate can simultaneously process 96 separate samples and contains a microporous membrane filter sealed to the underside of a standard microtiter plate. The plate is designed for use with the Millititer Vacuum Holder which is connected to a vacuum source. The liquid in each sample is removed by vacuum and the oligo dT-bound RNA is trapped in the wells. The samples in individual wells were washed with 500 μl (5×100 μl) of oligo dT loading buffer followed successively by 2.5 ml of loading buffer containing 0.1% SDS and 500 μl of loading buffer without SDS. The bound RNA was then eluted with 300 μl (3×100 μl) of water and collected in individual wells of a standard 96 well microtiter plate placed underneath the filtration plate. The eluted RNA was transferred to vials for scintillation counting or ethanol precipitated for further analysis. The amount of methyl incorporation in the different substrates was determined from the ^{3}H-DPM.

Analysis of methylated RNA

HPLC analysis was performed as described previously (6, 12) to confirm the presence of m^{6}A in the RNA substrates. T1 oligonucleotide analysis of unmethylated and methylated bPRL RNA was performed as follows. The bPRL RNA (^{32}P or ^{3}H labeled) along with 10 μg of yeast tRNA as carrier was digested with 2 μg of T1 RNase in 10 mM Tris HCl pH 7.5, 1 mM EDTA in a final volume of 4 μl at 37°C for 45 minutes (16). The resulting T1 oligonucleotides were separated by electrophoresis on a 20% polyacrylamide/7M urea gel, transferred to Zetaprobe membrane (Bio-rad), sprayed with En^{3}Hance (Dupont) and fluorographed. Analysis of mutant RNA (−1G to A) required an additional digestion with RNase A because both AAC sequences were present in the same 27 base T1 oligonucleotide. After electrophoresis of the T1 oligonucleotides, the 27 base oligonucleotide was identified by autoradiography and then eluted from the gel by standard techniques (27). Each oligonucleotide, along with 5 μg yeast tRNA as carrier, was digested with 2 μg RNase A at 37°C for 45 minutes and the resulting oligonucleotides were again separated by electrophoresis on a 20% denaturing polyacrylamide gel, transferred to Zetaprobe membrane, sprayed with En^{3}Hance and flourographed.

Quantification of methylation in the wild type 3′ terminal bPRL 60 RNA, the 5′ terminal bPRL 192 RNA and their mutated sequences was performed as follows: ^{32}P-GTP-labeled RNA was methylated *in vitro* as described above. The methylated RNA was digested with RNase T1 and the T1 oligonucleotideswere separated on a 20% denaturing polyacrylamide gel. The T1 oligonucleotides were transferred to a Zetaprobe membrane and autoradiographed. The bands containing the GAC and AAC sequences were cut out of the membrane and the radioactivity in the bands determined by scintillation counting. The moles of [^{3}H]methyl incorporated into each oligonucleotide were calculated from the ^{3}H-DPM and the amount of oligonucleotide recovered was determined by the ^{32}P-DPM. From these values the amount of methyl group incorporated per mole of RNA was calculated. This value was used to determine the relative levels of methylation in the wild type and mutant RNAs. To determine the level of methylation in the −1G to A mutant containing two AAC sequences, ^{32}P-ATP-labeled substrate RNA was methylated *in vitro*, digested with T1 RNase and the resulting 27-base T1 oligonucleotide containing both AAC sequences was eluted from the gel (27). The eluted 27-base T1 oligonucleotide was subsequently digested with RNase A, the products were separated on a 20% denaturing gel and transferred to Zetaprobe membrane. After autoradiography, the bands containing the AAC sequences were cut out, counted and the moles of methyl in each oligonucleotide were calculated as described above.

Methylation of double stranded RNA substrates

An RNA oligonucleotide complementary to that portion of bPRL 60 RNA containing the wild-type site of methylation was generously provided by United States Biochemical Corporation. The sequence of this 20-base oligonucleotide was CUACAGA-GUCUGUUUUUAUU. Conditions resulting in the optimal annealing of the oligonucleotide to bPRL RNA were monitored by protection of T1 RNase sensitive sites in the duplex to digestion by this enzyme. The oligonucleotide and bPRL 60 RNA were mixed in a volume of 10 μl of water and heated to 90°C for 1 minute. The reaction mix (30 μl) containing 20 mM Hepes, 80 mM KCl, 1.7 mM $MgCl_2$, 3% PVA and 2μM [^{3}H]-SAM was added immediately and the tubes incubated at 30°C for 5 minutes. HeLa cell nuclear extract (10 μl) was then added and the reaction allowed to proceed for either 10 minutes or 30 minutes at 30°C. The bPRL RNA was purified by binding to oligo dT cellulose as described above. To generate a short non-complementary RNA that would serve as a control for these reactions, the vector pBSM13+ was digested with Bam H1 and a 40 nucleotide unlabeled RNA sequence was synthesized from the T3 RNA polymerase promoter. The RNA (T3 RNA) was separated from the unincorporated nucleotides by passing the reaction mixture over a G-25 gel filtration column and the amount of T3 RNA synthesized was determined by absorbance at 260 nM. A second non-specific RNA was also synthesized in a similar manner. This sequence is 90 nucleotides long and contains the splice leader sequence (SL RNA) from *Schistosoma mansoni* (28). The T3 and SL RNA replaced the RNA oligonucleotide in the control reactions.

RESULTS

Methylation of the 3′ terminus of prolactin mRNA

We previously demonstrated that the bPRL mRNA is methylated predominantly at a single adenosine, both *in vivo* and *in vitro* (16). This methylation site is located in the AGACU sequence found in the 3′ untranslated region of the mRNA. To determine if only a portion of the 3′ bPRL mRNA containing this consensus sequence was sufficient for accurate methylation we synthesized an RNA transcript (bPRL 60) containing 60 nt of the 3′ end of bPRL mRNA followed by 50 A residues (Figure 1A) and determined its methylation *in vitro*. We also utilized this transcript to develop an assay that allowed the substrate RNA to be easily purified from the methylation reaction by binding to oligo dT cellulose (see Materials and Methods). HPLC analysis of the methylated RNA indicated that the only modification (90% of input DPM) in the RNA was m^{6}A (data not shown). Accuracy of methylation was determined by RNase T1 oligonucleotide analysis (Fig. 2A). The T1 oligonucleotides from ^{32}P-GTP-labeled unmethylated RNA is shown in lane a (compare with T1 oligonucleotides shown in Fig. 1A). Lanes b and c show only those T1 oligonucleotides that contain a methylated adenosine. Nearly all of the [^{3}H]methyl radioactivity was associated with

Figure 1. **A.** The sequence of the 3′ 60 nucleotides of bPRL mRNA (bPRL 60) linked to 50 adenosine residues at the 3′ end. The T1 oligonucleotides are boxed and their respective sizes are indicated. The oligonucleotides containing the AAC and GAC consensus sequences are 21 and 6 nucleotides long, respectively, and are shaded. The adenosines that can be methylated are in bold print. **B.** The sequence of the 5′ 192 nucleotides of bPRL mRNA (bPRL 192) linked to 50 adenosine residues. The T1 oligonucleotides are boxed and their sizes are indicated. The oligonucleotides containing the AAC and GAC consensus sequences are shaded and the adenosines that can be methylated are in bold print.

the AGACU sequence (6-base oligonucleotide) while the AAC sequence (21-base oligonucleotide) located one nucleotide upstream was poorly methylated. The presence or absence of 7-methyl cap structures at the 5′ end did not influence methylation of internal adenosine residues (data not shown) and all experiments reported here were performed with uncapped RNA. Removal of the poly A tail from the 3′ end of the bPRL substrate by oligo dT directed RNase H cleavage did not affect methylation of the GAC sequence (data not shown).

Effect of mutations in the consensus sequence on methylation

The presence of a single primary methylation site in a naturally occuring mRNA makes the prolactin mRNA a good substrate for the analysis of elements involved in methylation site recognition. Although RGACU sequences are clearly required for accurate methylation, they are not sufficient for defining the sites of methylation. For example, two other RGACU sequences present in bPRL mRNA are poor substrates for the methyltransferase. To gain further insight into the nature of sequence elements (consensus and nonconsensus) which are important for methylation site definition we made site-specific mutations within the -AAACAGACU- sequence of bPRL 60 and tested these mutants for accurate and efficient methylation in the HeLa cell nuclear extract. These mutants also allowed us to directly determine the preferred extended consensus sequence for methylation at the AAC sequence. Methylation at the predominant GAC (site 1) and AAC (site 2) sequences within the -AAAC-AGACU- region were identified by T1 oligonucleotide analysis. The adenosine residue which is methylated in the sequence A-GACU is denoted 0. Nucleotides downstream of the adenosine are referred to as +1, +2 while nucleotides upstream of the adenosine are denoted −1, −2 (Fig. 2B). The levels of methylation in the mutant and wild type sequences at limiting RNA concentrations were determined as described in Materials and Methods and were the average of at least 3 determinations. Methylation of the GAC sequence (site 1) in wild type bPRL RNA was assigned a value of one and methylation at the AAC

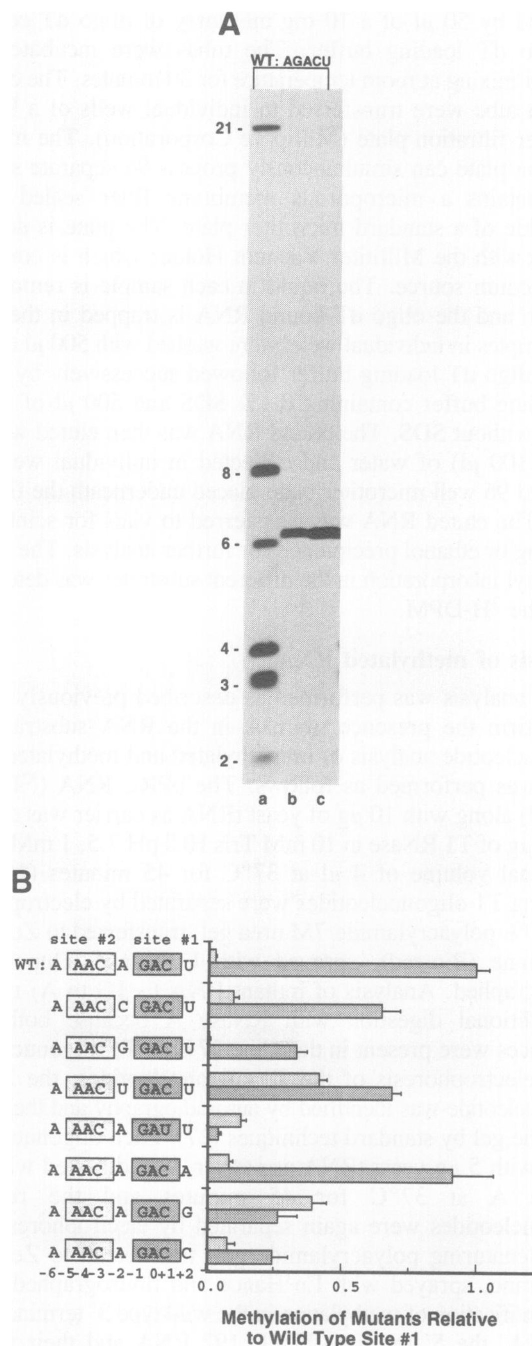
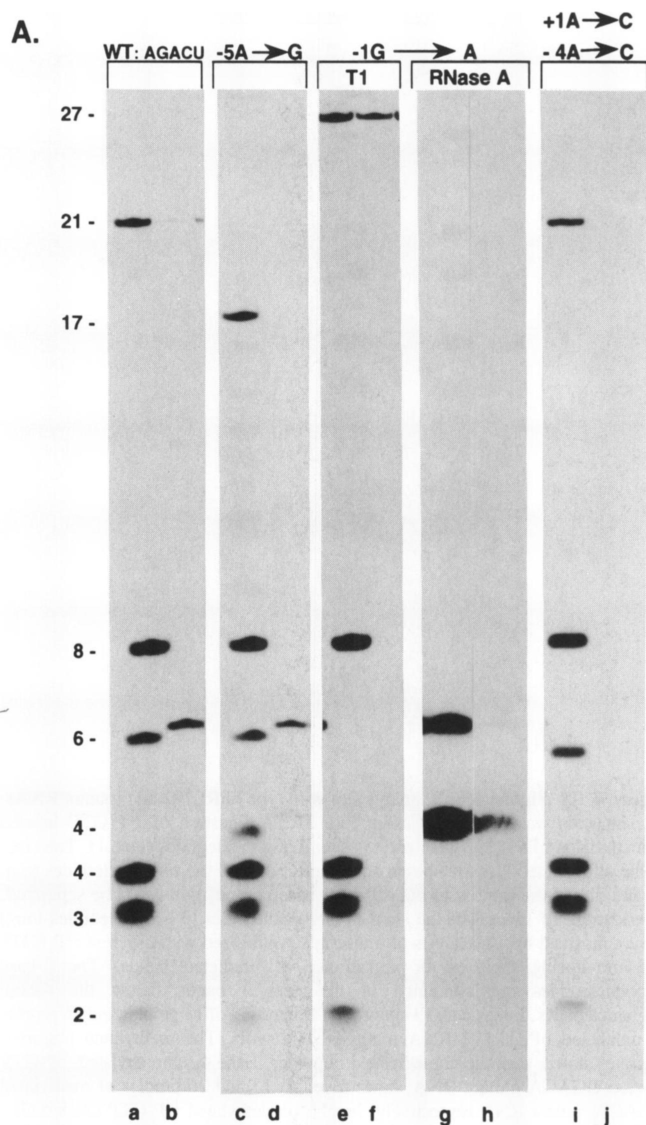
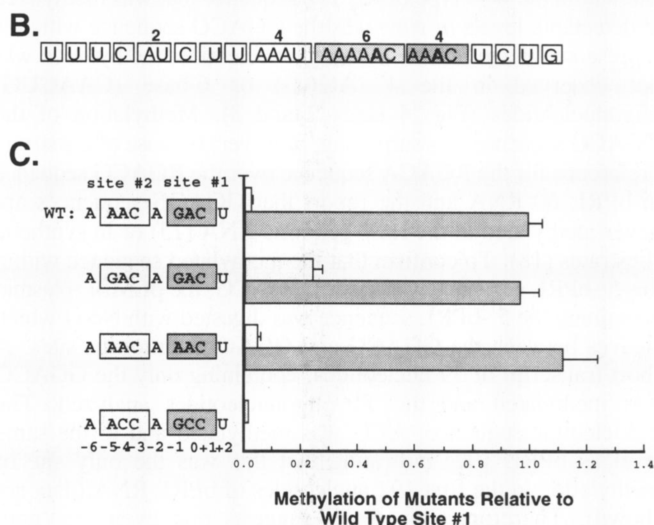


Figure 2. **A.** Oligonucleotide analysis of wild type bPRL 60 RNAs. Unmethylated ^{32}P-GTP labeled RNA (lane a) and unlabeled RNA methylated *in vitro* (lanes b and c) were digested with RNase T1, the oligonucleotides were separated by denaturing gel electrophoresis, transferred to Zetaprobe membrane and flourographed. The unmethylated RNA gave rise to several oligonucleotides whose sizes are indicated on the left. The 59-base oligonucleotide containing the poly A tail is present near the wells at the top of the gel and is not shown in this figure. Only the 6-base oligonucleotide containing the [^{3}H]methyl label is seen in lanes b and c. **B.** Analysis of mutations in the primary methylation site of bPRL 60 RNA. The location of the methylated adenosine in wild type bPRL 60 is indicated as 0. Methylation at both the GAC (site 1) and AAC (site 2) sequences were quantified by T1 oligonucleotide analysis as described in Methods and are expressed relative to the GAC methylation of wild type RNA (value of 1). Numbers are the mean of at least 5 determinations and the error bars represent the standard deviation of the mean.





sequence (site 2) is expressed relative to the level of GAC methylation. Methylation of all mutant RNAs (at both AAC and GAC sites) are expressed relative to the wild type GAC (site 1) methylation.

Mutation of the −2A of the AGACU sequence to C,G or U decreased methylation at this site (Fig. 2B). In constrast to the RSV genomic RNA, GGACU is not the preferred sequence in bPRL mRNA. The −2A mutants also affected methylation of the AAC sequence in site 2; −2A to G of AGACU almost completely inhibited methylation while −2A to U increased methylation 13-fold at site 2. Therefore, in this context, the core AAC sequence strongly prefers U at the −2 position over C or A while G is inhibitory.

Mutation of +1C to U in the AGACU sequence caused a 96% inhibition at site 1, confirming that the C residue of the core consensus is essential for methylation. Mutation of +2U to A at site 1 did not alter GAC methylation, but mutation of +2U to G or C decreased GAC methylation to 25% of wild type levels. The relative levels of methylation of these mutants were also examined at 10 min when methylation is approximately at 50% completion and were found to be the same as at 30 min (data not shown). All of these down mutations in the +1 and +2 position of the AGACU sequence (site 1) resulted in an increase in the level of methylation in AAC (site 2) with the +2U to G mutation at site 1 causing a 12-fold increase in absolute levels of AAC methylation compared to the wild type RNA. These results suggest that the GAC and AAC sequences may compete with each other for the methyltransferase; a reduction in methylation at the stronger GAC site can increase methylation at the weaker AAC site, or that sequences beyond the 5 base consensus (context) may influence methylation.

Context effects on methylation

To further explore this interaction between the GAC and AAC sites of methylation, additional mutants containing two GAC or two AAC sites were constructed. Mutation of −5A of -AAAC-AGACU- to G results in a substrate with two adjacent GAC

Figure 3. A. Oligonucleotide analysis of wild type bPRL 60 and mutant RNAs. Unmethylated ^{32}P-GTP labeled RNA (lanes a, c and i) and unlabeled RNA methylated *in vitro* (lanes b, d and j) were digested with RNase T1, the oligonucleotides were separated by denaturing gel electrophoresis, transferred to Zetaprobe membrane and flourographed. The unmethylated RNA gave rise to several oligonucleotides whose sizes are indicated on the left. The 59-base oligonucleotide containing the poly A tail is present near the wells at the top of the gel and is not shown in this figure. Only those oligonucleotides containing the [^{3}H]methyl label are seen in lanes b, d and j. The T1 and RNase A analysis of the −1G to A mutant is shown in lanes e−h. T1 RNase digestion of unmethylated ^{32}P-ATP labeled RNA (lane e); T1 RNase digestion of 13 pmoles of unlabeled RNA methylated *in vitro* (lane f); RNase A digestion of the 27 base T1 oligonucleotide from unmethylated RNA (lane g); RNase A digestion of 27 base T1 oligonucleotide from methylated RNA (lane h). **B.** Sequence of the 27 base T1 oligonucleotide generated by the −1G to A mutant. The oligonucleotides resulting from digestion with RNase A are boxed and their sizes are indicated. The oligonucleotides containing the AAC consensus sequences are shaded. Site 1 is present in the 4-base oligonucleotide and site 2 is present in the 6-base oligonucleotide. The adenosine that can be methylated in each of these sequences is in bold print. **C.** Quantitative analysis of mutants shown in Figure 3A. Methylation at site 1 and 2 of the mutant RNAs is expressed relative to methylation at the GAC (site 1) sequence of wild type bPRL 60 (value of 1). Numbers are the mean of at least 5 determinations and the error bars represent the standard deviation of the mean.

sequences. The upstream (site 2) AGACA is found in the 4-base T1 oligonucleotide and was methylated to only 25% of the downstream (site 1) AGACU wild type sequence present in the 6-base T1 oligonucleotide (Fig 3A lane d and Fig. 3C). However, when an AGACA sequence is created at site 1 (+2U to A mutation), it is methylated as well as the strong consensus A-GACU sequence normally found at this position (Fig. 2B), suggesting that context effects can alter the efficiency of methylation of consensus sequences. Mutation of the −1G to A in site1 of -AAACAGACU- creates two adjacent AAC sequences; AAACA present upstream (site 2) and AAACU sequence present downstream (site 1). Both of these sequences are present in a single 27-base T1 oligonucleotide (Fig. 3A lanes e and f and Fig.3B) and resolution between the two sites required an additional approach. To differentiate between the two AAC sequences, the 27-base T1 oligonucleotide labeled with ^{32}P-ATP (Fig. 3A lane e) was purified from the gel and digested with RNase A to produce two 4-base (AAAU and AAAC) and one 6-base (AAAAAC) oligonucleotides (Fig. 3A lane g and Fig. 3B). The site 2, AAACA, is present in the 6-base oligonucleotide and the site 1, AAACU, is present in the 4-base oligonucleotide. The T1 oligonucleotide from methylated RNA (Fig. 3A lane f) was also digested with RNase A (Fig. 3A lane h). Methylation occurs predominantly at the site 1 AAACU sequence (Fig. 3A lane h). As shown in Fig. 3C, this AAACU sequence at site 1 is now methylated as well as the AGACU sequence normally present at this site in the wild type bPRL 60 RNA. This was a surprising observation because even though AAACU is a much better substrate than AAACA, it is not methylated as well as an AGACU sequence when present at site 2 (Fig 2B, −2A to U mutant). These results demonstrate that an identical sequence can be methylated to strikingly different levels, depending on its context within the mRNA.

Finally, a mutant was constructed in which the middle adenosine in both the AAC and GAC sequences in 3′ bPRL60 were mutated to C (Fig. 3C). The mutated GAC sequence cannot be methylated because the adenosine is no longer present. However, these changes fortuitously create a new AAC sequence; the −5A can now be potentially methylated. Methylation at this sequence (present in the 21-base T1 oligonucleotide)is undetectable (Fig. 3A, lane j and Fig. 3C), even though the sequence AAACC when present in other locations of the mutated bPRL substrate is methylated better than the wild type AAACA sequence (Fig. 2B). This observation further underscores the importance of the immediate sequence context in defining methylation sites.

To further document the influence of consensus sequence location or context effects on the efficiency of methylation, we examined methylation of both naturally occuring and mutated consensus sites within the 5′ region of prolactin RNA. The 5′ end of prolactin is not methylated at detectable levels *in vivo* (11) and only methylated at low levels *in vitro* (16). Examination of the sequence of the first 192 nt of the 5′ end of prolactin RNA indicated that it contains three consensus sequences GGACG, GGACA and CAACU (Fig 1B). An RNA transcript containing 192 nucleotides of the 5′ end of prolactin RNA and 50 terminal A residues (bPRL 192) was synthesized and its methylation studied *in vitro*. HPLC analysis of the albeit poorly methylated wild type 192 RNA confirmed the presence of only m^{6}A (data not shown). T1 oligonucleotide analysis of the methylated RNA was performed to determine the site(s) of methylation. As indicated in Fig. 1B the T1 oligonucleotides containing the three

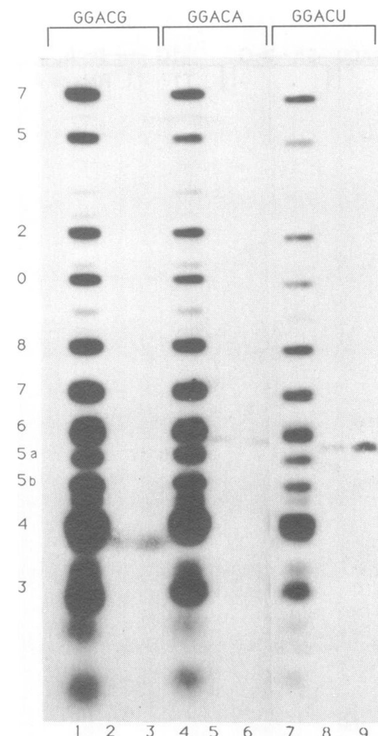


Figure 4. T1 oligonucleotide analysis of wild type bPRL192 and mutant RNAs. The analysis was performed as in Fig. 3. T1 digestion of ^{32}P-GTP labeled unmethylated bPRL 192 gave rise to several oligonucleotides (lane 1). The sizes of the oligonucleotides are shown on the left. Some of the oligonucleotides (e.g. 5a and 5b) of the same size but different base composition could be separated. The identity of several of the smaller oligonucleotides (3−5 nucleotides long) were confirmed by T1 analysis of transcripts synthesized with different ^{32}P-NTPs and monitoring the loss or appearance of predicted bands. The 3-base oligonucleotides are contained in the next 3 bands below the 4-base oligonucleotides. Lanes 2 and 3 show the T1 digestion of 13 pmoles and 26 pmoles of unlabeled bPRL 192 RNA methylated *in vitro*. The methylated oligomer migrates slower than the unmethylated oligomer. Lane 4, unmethylated ^{32}P-GTP labeled GGACA mutant RNA; lanes 5 and 6, 13 and 26 pmoles of methylated GGACA mutant RNA respectively; lane 7, unmethylated ^{32}P-GTP labeled GG-ACU mutant RNA, lanes 8 and 9, 13 and 26 pmoles of methylated GGACU mutant RNA, respectively.

consensus sequences are ACG, ACAG and CAACUG. The only site within the wild type bPRL 192 sequence that was methylated at detectable levels *in vitro* was the GGACG sequence which is represented in the 3-base ACG oligonucleotide; methylation was not observed in the 4 (ACAG) or 6-base (CAACUG) oligonucleotides (Fig. 4 lanes 2 and 3). Methylation of the GGACG sequence was surprising, however, because of the strong preference for the RGACA sequence over the RGACG sequence in bPRL 60 RNA and the report that GGACG sequences are never methylated in the RSV genomic RNA (15) or in synthetic substrates (18). To confirm that the methylated sequence within this 5′-bPRL sequence was indeed GGACG, the pEMBL plasmid containing the 5′-bPRL sequence was digested with Nco I which cleaves between the GGACG and GGACA consensus sites. A short transcript of 59 nucleotides, containing only the GGACG was methylated and the T1 oligonucleotides analyzed. The trinucleotide sequence, ACG was methylated and to the same level as bPRL 192, indicating that this was the only site of methylation in the first 192 nucleotides of bPRL RNA (data not shown). Therefore, this result suggests that even the 'non-consensus' GGACG sequence nonetheless can be methylated,

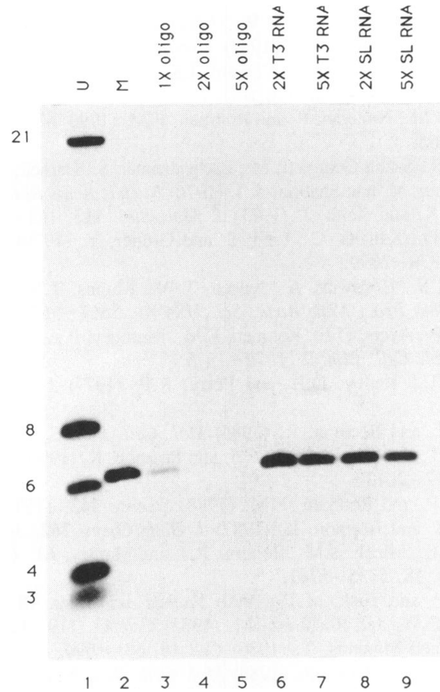


Figure 5. Effect of double stranded RNA substrate on methylation. Unlabeled bPRL 60 RNA was methylated in the absence or presence of the complementary RNA oligonucleotide. Methylation at the GAC sequence was determined by T1 oligonucleotide analysis. Only the methylated oliogonucleotides contain ^3H label which was visualized by flourography. The sizes of the oligonucleotides are shown on the left. Lane 1, unmethylated ^{32}P-GTP labeled bPRL 60 RNA ; lane 2, bPRL 60 RNA methylated in the absence of complimentary RNA; lanes 3–5, bPRL 60 RNA methylated in the presence of equal, 2 or 5 fold molar excess of complementary RNA; lanes 6–9, methylated bPRL 60 RNA in the presence of 2 or 5 fold molar excess of nonspecific T3 or SL RNA.

depending on the RNA context within which it resides. When this GGACG sequence was mutated to either GGACA or GGACU, the GGACA methylation was decreased by 70% (Fig. 4 lanes 5 and 6) whereas the GGACU methylation increased 2-fold (Fig. 4 lanes 8 and 9). Comparison of the relative levels of methylation of the GGACU mutant in 5′ bPRL 192 with the GGACU mutant in 3′ bPRL 60 indicated that methylation of the GGACU in 5′ bPRL was 3-fold greater than the same sequence at the 3′ end, again consistent with the observation that both primary sequence and its context in the mRNA play a role in defining methylation sites.

Effect of RNA duplex on methylation

The observation that m^{6}A methylation of prolactin mRNA occurs predominantly at the 3′ end of the molecule, which is significantly enriched in A+U content along with the context dependency of consensus sequence utilization suggested that a single stranded region may be required for efficient methylation. To test this, we hybridized bPRL 60 RNA to an RNA oligonucleotide complementary to a 20 nucleotide region of the 3′ end of prolactin RNA containing the AAC and GAC methylation consensus sequences. The formation of the duplex was monitered by T1 RNase digestion; sites within the duplex were insensitive to T1 RNase digestion (data not shown). The duplexed RNA was incubated in the methylation reaction, the substrate RNA was purified from the reaction and T1 oligonucleotide analysis was performed. Methylation at the GAC (6-base oligonucleotide) was inhibited 80–90% with an equimolar amount of the complementary oligonucleotide (Fig. 5 lane 3) nor was methylation detected at the AAC site (21-base oligonucleotide). With a 5-fold molar excess of the complementary oligonucleotide, GAC methylation was completely inhibited (Fig. 5 lane 5). In contrast, the non-complementary RNA sequences, T3 RNA or SL RNA, had little effect on methylation (Fig. 5 lanes 6–9). A five fold molar excess of T3 RNA or SL RNA resulted in only slight inhibition, presumably in a non-specific manner. Clearly, a long, stable duplex structure is not a substrate for the N^6-adenosine methyltransferase; however, the role of more subtle secondary structures present in mRNA remains to be determined.

DISCUSSION

Utilization of wild type and mutant sequences of bPRL mRNA allowed the characterization of consensus m^{6}A methylation sites in the context of a natural mRNA sequence and has revealed several features of consensus site recognition that differ significantly from previous reports. Although AGACU/A was the most active methylation sequence, GGACU, which also contains a purine at −2, was only a poor substrate for the methyltransferase. In both RSV genomic RNA (15) and in experiments utilizing a synthetic 20 nucleotide substrate RNA containing multiple copies of a m^{6}A methylation site (18) GGACU was the sequence of choice while GGACG is not methylated. Depending on sequence context, however, GGACG can be methylated in bPRL mRNA (Fig. 4).

Previous studies on m^{6}A methyltransferase sequence specificity have focused on the GAC sequence because it is more commonly methylated than AAC sequences; the requirement for an extended AAC consensus sequence had not been previously tested. The results presented in this study demonstrate the existence of an extended AAC consensus sequence and in the context tested (bPRL mRNA) the AAACU sequence is strongly preferred over the AAACA/C/G sequences. The differences in sequence specificity observed between the bPRL, RSV, and synthetic RNAs suggest that the sequence constraints vary in different RNA substrates. For example, though the RSV genomic RNA does not contain any methylated AAC sites, we find active AAC sites in the last intron of bPRL pre-mRNA (7) and can create AAC consensus sequences whose efficiencies rival those of GAC sequences. Though it is certain that the core AAC and GAC sequences are essential for m^{6}A mRNA methylation, the enzyme apparently tolerates a much greater degeneracy in the extended 5 nucleotide consensus sequence than previously thought possible. In this respect, m^{6}A methylation parallels other RNA–protein recognition systems such as splice donor/acceptor consensus sites which are composed of both highly conserved and degenerate nucleotides (29). It is important to recognize that the precise sites of methylation have been determined only in the RSV genomic RNA and in bPRL mRNA. Methylation sites in additional mRNAs must be defined before generalizations concerning the contributions of the various bases in the extended consensus sequence can be determined.

Utilizing bPRL mRNA, we have demonstrated that a poorly methylated sequence can be converted into a sequence that is methylated at reasonably high efficiency merely by changing its location within the larger RNA context. Mutational analysis of adjacent AAC or GAC sequences within the 3′-bPRL mRNA has shown that identical sequences can be methylated to different

levels, depending on the immediate context of the methylation site. Therefore, when competing consensus sequences are present in any RNA molecule, the information provided by surrounding sequences (methylation site context) may be as important as the primary sequence of the 5 nucleotide consensus in defining the preferred sites of methylation. Such context effects have also been observed to influence the splicing efficiency and selection of splice sites in yeast and mammals (20–24). Unfortunately it is not yet possible to identify the specific interactions responsible for the context effects on mRNA methylation. Though RNA secondary structure might be involved in determining the relative efficiency of two widely separated sequences, it is unlikely to affect substrates with single base changes at closely adjacent methylation sites such as site 1 and site 2 of the -AAACAGACU- sequence. In this regard, computer analysis of the structure of several single nucleotide mutants which varied significantly as methylation substrates did not generate any detectable differences in the predicted secondary structure.

Our experiments with complementary RNA sequences indicate that strong methylation sites are not recognized when present in stable, duplexed regions. If RNA molecules exist in multiple conformations it is possible that only the fraction of RNA molecules containing the consensus sequence in a region of non-duplexed structure are capable of being methylated. This would offer a possible explanation for the less than stoichiometric levels of methylation observed in both the RSV and PRL RNAs (14–16).

While these studies offer an understanding of the contribution of consensus sequence and context effects to recognition of the RNA substrate, the underlying mechanism of the reaction remains poorly understood. Although a consensus sequence is required for methylation, we do not know if the enzyme binds at or near the consensus sequence. It is also possible that separate enzyme activities catalyze the m^6A methylation of AAC and GAC sequences. The efficiency of methylation is a combination of both the binding affinity of the enzyme to the RNA substrate and the rate of catalysis. Unfortunately, these components cannot be resolved in crude nuclear systems. It will be of interest to determine how mutant substrates affect each step of the reaction once the methyltransferase is purified. Our detailed analysis of methylation within a segment of a naturally occurring mRNA indicates that factors in addition to a simple 5 nucleotide consensus sequence play a role in the recogniton and efficient methylation of mRNA substrates by N^6-adenosine methyltransferase.

ACKNOWLEDGEMENTS

We gratefully acknowledge Josh Bloom and Karl Schwenk for help in the early stages of this project, Dave Ayers for nuclear extract preparation and Tom Kienzle for generously providing the SL RNA. We thank Joe Bokar, John Nilson, Mary Eileen Rath, and Tom Kienzle for helpful comments on the manuscript. E.C.G. was supported by a National Science Foundation Graduate Fellowship. This research was supported by NIH grant CA 31810 from the National Cancer Institute to F.M.R.

REFERENCES

1. Banerjee, A.K. (1980) *Microbiol. Review* **44**, 175–205.
2. Narayan, P. and Rottman, F.M. (1992) *Adv. in Enzymology* **65**, 255–286.
3. Bachellerie, J.P., Amalric, F. and Caboche, M. (1978) *Nucleic Acids Res.* **5**, 2927–2943.
4. Stoltzfus, C.M. and Dane, R.W. (1982) *J. Virol.* **42**, 918–931.
5. Finkel, D. and Groner, Y. (1983) *Virology* **131**, 409–425.
6. Camper, S.A., Albers, R.J., Coward, J.K. and Rottman, F.M. (1984) *Mol. Cell. Biol.* **4**, 538–543.
7. Carroll, S.M., Narayan, P. and Rottman, F.M. (1990) *Mol. Cell. Biol.* **10**, 4456–4465.
8. Sommer, S., Salditt-Georgieff, M., Bachenheimer, S., Darnell, J.E., Furuichi, Y., Morgan, M. and Shatkin, A.J. (1976) *Nucleic Acids Res.* **3**, 749–765.
9. Beemon, K. and Keith, J. (1977) *J. Mol. Biol.* **113**, 165–179.
10. Canaani, D., Kahana, C., Lavi, S. and Groner, Y. (1979) *Nucleic Acids Res.* **6**, 2879–2899.
11. Horowitz, S., Horowitz, A., Nilsen, T.W., Munns, T.W. and Rottman, F.M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5667–5671.
12. Narayan, P., Ayers, D.F., Rottman, F.M., Maroney, P.A. and Nilsen, T.W. (1987) *Mol. Cell. Biol.* **7**, 1572–1575.
13. Schibler, U., Kelley, D.E. and Perry, R.P. (1977) *J. Mol. Biol.* **115**, 695–714.
14. Kane, S.E. and Beemon, K. (1985) *Mol. Cell. Biol.* **5**, 2298–2306.
15. Csepany, T., Lin, A., Baldick, C.J. and Beemon, K. (1990) *J. Biol. Chem.* **265**, 20117–20122.
16. Narayan, P. and Rottman, F.M. (1988) *Science* **242**, 1159–1162.
17. Kane, S.E. and Beemon, K. (1987) *J. Biol. Chem.* **262**, 3422–3427.
18. Harper, J.E., Miceli, S.M., Roberts, R.J. and Manley, J.L. (1990) *Nucleic Acids Res.* **18**, 5735–5741.
19. Rana, A.P. and Tuck, M.T. (1990) *Nucleic Acids Res.* **18**, 4803–4807.
20. Pikielny, C.W. and Roshbash, M. (1985) *Cell* **41**, 119–126.
21. Reed, R. and Maniatis, T. (1986) *Cell* **46**, 681–690.
22. Furdon, P.J. and Kole, R. (1988) *Mol. Cell. Biol.* **8**, 860–866.
23. Nelson, K.K. and Green, M.R. (1988) *Genes Dev.* **2**, 319–329.
24. Goguel, V. and Roshbash, M. (1993) *Cell* **72**, 893–901.
25. Setzer, D.R., Hmiel, R.M. and Liao, S. (1990) *Nucleic Acids Res.* **18**, 4175–4178.
26. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
27. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor.
28. Rajkovic, A., Davis, R.E., Simonsen, J.N. and Rottman, F.M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8879–8883.
29. Green, M.R. (1991) *Annu. Rev. Cell Biol.* **7**, 559–599.