Efficiency of N-Linked Core Glycosylation at Asparagine-319 of Rabies Virus Glycoprotein Is Altered by Deletions C-Terminal to the Glycosylation Sequon†

Susan H. Shakin-Eshleman,‡ William H. Wunner,§ and Steven L. Spitalnik*,‡

University of Pennsylvania School of Medicine and The Wistar Institute, Philadelphia, Pennsylvania 19104

Received March 16, 1993; Revised Manuscript Received June 3, 1993®

ABSTRACT: In N-linked core glycosylation, the oligosaccharide $Glc_3Man_9GlcNAc_2$ is transferred to the tripeptide sequon Asn-X-Ser/Thr. However, this process must be regulated by additional protein signals, since many sequons are either poorly glycosylated or not glycosylated at all. Since N-linked glycosylation can influence protein structure and function, understanding these signals is essential for the design and expression of recombinant glycoproteins. Core glycosylation usually occurs cotranslationally in the rough endoplasmic reticulum (RER) during translocation of nascent proteins. Since only regions of a protein immediately near to a sequon or N–terminal to it are thought to be in the RER when core glycosylation occurs, most models predict that regions C-terminal to the sequon do not influence this process. We tested whether regions C-terminal to a sequon can influence its core glycosylation. Full-length (505 amino acid) rabies virus glycoprotein (RGP) mutants, each containing only one of the three sequons normally present in RGP, were used for these studies. Using a cell-free system, the core glycosylation efficiency at each sequon was determined. Termination codons were then introduced into these mutants at defined sites to produce C-terminal truncations, and the effect of each of these truncations on the core glycosylation efficiency at each sequon was assessed. While deletion of the C-terminal transmembrane and cytoplasmic domains did not affect core glycosylation, more extensive C-terminal deletions did result in altered core glycosylation in a site-specific fashion. Specifically, C-terminal truncations resulting in proteins containing 386 or 344 amino acids decreased the efficiency of core glycosylation at Asn319. This demonstrates that core glycosylation efficiency can be influenced by the presence or absence of regions in a protein more than 68 amino acids C-terminal to a specific glycosylation site.

Many proteins are modified by the addition of oligosaccharides to specific asparagine (Asn) residues [for review, see Kornfeld and Kornfeld (1985)]. This process, known as N-linked glycosylation, can influence many properties of proteins, including intracellular transport (Machamer & Rose, 1988; Dube et al., 1988; Semenkovich et al., 1990; Ng et al., 1990), biological activity (Dube et al., 1988; Semenkovich et al., 1990; Tao & Morrison, 1989; Matzuk et al., 1989), stability (Matzuk & Boime, 1988; Pizer et al., 1980), and antigenicity (Wright et al., 1989; Klenk, 1990; Hobman et al., 1991). The biological effects of N-linked glycosylation often depend on the site of glycosylation within a polypeptide chain (Machamer & Rose, 1988; Dube et al., 1988; Ng et al., 1990; Matzuk et al., 1989). In addition, inefficient glycosylation at a specific site can lead to the synthesis of alternate glycoforms of a protein which differ from one another only by the presence or absence of an oligosaccharide at a specific site (Plummer & Hirs, 1964; Wunner et al., 1985); these glycoforms may differ from one another with regard to biological function. Therefore, complete understanding of the regulation of glycoprotein synthesis and function requires the characterization of the protein signals which regulate the efficiency of N-linked glycosylation at specific sites.

The presence or absence of N-linked oligosaccharides at specific sites in a glycoprotein is determined during the initial step of N-linked glycosylation, known as core glycosylation. Core glycosylation is believed to occur cotranslationally as nascent proteins are translocated into the lumen of the rough endoplasmic reticulum (RER).[1] In this step, the presynthesized core oligosaccharide, $Glc_3Man_9GlcNAc_2$, is transferred from the lipid carrier dolichol to an Asn residue in the nascent peptide chain by the enzyme oligosaccharyltransferase [for review, see Kornfeld and Kornfeld (1985) and Kaplan et al. (1987)]. Core glycosylation generally occurs only at Asn residues in the sequon Asn-X-Ser/Thr, where X is any amino acid except proline (Kaplan et al., 1987; Marshall, 1974). However, while this sequon is required for N-linked glycosylation, other protein signals must also influence core glycosylation of specific Asn residues, since many Asn-X-Ser/Thr sequons are either inefficiently glycosylated or not glycosylated at all in vivo (Plummer & Hirs, 1964; Wunner et al., 1985; Mononen & Karjalainen, 1984; Nakai & Kanehisa, 1988).

Current understanding of the protein signals which influence core glycosylation efficiency comes predominantly from studies using membrane extracts as a source of oligosaccharyltransferase to glycosylate synthetic peptides or proteolytic fragments (Hart et al., 1979; Struck et al., 1978; Kronquist & Lennarz, 1978). Such studies have identified several features of the primary structure of peptides near Asn residues which influence their activities as oligosaccharide acceptors [for review, see

† S.H.S.-E. is supported by National Research Service Award 1 F32 GM 14373-01 from the National Institutes of Health and by the MetPath Fellowship in Laboratory Medicine. S.L.S. is a Mallinckrodt Scholar.

* To whom correspondence should be addressed at the Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, 220 John Morgan Building, 36th St. and Hamilton Walk, Philadelphia, PA 19104. Phone: (215) 898-8063; Fax: (215) 898-4227.

‡ University of Pennsylvania School of Medicine.

§ The Wistar Institute.

® Abstract published in *Advance ACS Abstracts,* August 15, 1993.

[1] Abbreviations: RGP, rabies virus glycoprotein; RER, rough endoplasmic reticulum; PAGE, polyacrylamide gel electrophoresis; PM, dog pancreatic microsomes.

0006-2960/93/0432-9465$04.00/0

Kaplan et al. (1987)]. These studies further suggested that the core glycosylation efficiency at specific sequons in vivo may be inhibited by protein folding, involving interactions with regions distant from the sequon. This is supported by the observation that some sequons which are not glycosylated in vivo can be glycosylated posttranslationally following denaturation or proteolytic degradation (Struck et al., 1978; Kronquist & Lennarz, 1978; Pless & Lennarz, 1977). However, studies of this kind have not defined specific regions of proteins that have the potential to influence core glycosylation at individual sequons.

While the peptide-based systems described above yielded considerable information about the transfer of the core oligosaccharide to peptide structures, studies performed using this experimental approach fail to address certain topological issues related to core glycosylation in vivo. Since core glycosylation typically occurs during translation and translocation, and since the protein translocation channel, the dolichol oligosaccharide donor, and oligosaccharyltransferase all reside within the RER membrane, the process of core glycosylation appears to be highly constrained, both spatially and temporally (Kaplan et al., 1987; Kelleher et al., 1992). Most models of core glycosylation propose that oligosaccharide transfer occurs immediately after a glycosylation sequon is translocated into the RER lumen (Kaplan et al., 1987; Kelleher et al., 1992; Rothman & Lodish, 1977; Glabe et al., 1980). In this case, only regions of a nascent protein very close to a sequon or upstream (N-terminal) from a sequon would be present within the RER lumen and available for protein folding interactions at the time when core glycosylation occurs. In contrast, regions of a nascent protein significantly downstream (C-terminal) from a sequon would be either sequestered in the RER membrane, sequestered within the large subunit of the ribosome, or not yet translated when the sequon is introduced into the RER lumen; these C-terminal regions would not be expected to influence the efficiency of core glycosylation at upstream sites (Finley et al., 1990).

In this report, we directly test whether regions of a protein significantly C-terminal to a sequon can influence the efficiency of core glycosylation. The model protein used in these studies, rabies virus glycoprotein (RGP), contains three sequons for N-linked glycosylation within the extracellular domain. In a previous study, we directly compared the efficiencies of core glycosylation at each of these three sequons using a series of RGP glycosylation mutants in which one or more of the sequons in RGP was deleted by site-directed mutagenesis (Shakin-Eshleman et al., 1992). Analysis of these mutants, both in a cell-free system and in transfected tissue culture cells, revealed efficient core glycosylation at Asn247 and Asn319, with inefficient core glycosylation at Asn37. In the present report, we use these RGP glycosylation mutants to test whether the core glycosylation efficiency of a sequon is influenced by regions of the protein significantly downstream (C-terminal) from the sequon. To do this, termination codons were introduced at defined sites C-terminal to the glycosylation sequons to truncate the previously described RGP glycosylation mutants. The effects of these deletions on the efficiency of cotranslational core glycosylation at each sequon in the RGP were then assessed using a cell-free transcription/translation/glycosylation system.

EXPERIMENTAL PROCEDURES

Construction of RGP Termination Mutants. The construction of plasmid pRGP(WT) which contains the cDNA for RGP (Kieny et al., 1984) inserted into the *Bgl*II site of pSG5 (Green et al., 1988) was previously described (Shakin-Eshleman et al., 1992). The RGP cDNA extends from position 1055 to 2706, where base 1068 corresponds to the A of the ATG initiation codon. The construction of plasmids encoding glycosylation mutants of RGP, pRGP(---), pRGP(1--), pRGP(-2-), and pRGP(--3), was previously described (Shakin-Eshleman et al., 1992).

Termination mutants containing one or more of the glycosylation sequons in RGP were generated by inserting the *Nhe*I amber stop linker, CTAGCTAGCTAG (Pharmacia), at selected restriction enzyme sites in the RGP cDNA of the plasmids described above using standard methods (Sambrook et al., 1989). The position of the termination codon in the resulting plasmids has been indicated by the prefix, T, followed by the amino acid position of the termination codon. The restriction sites used for linker insertion were as follows: T345, *Eco*NI (position 2142); T387, *Nde*I (position 2274; T434, *Hin*cII (position 2412). In these mutants, translation of the stop linker sequence prior to termination directs the synthesis of the following C-terminal amino acids: T345, Leu342Ala343Ser344; T387, Thr385Ser386; T434, Leu431Ala432Ser433. Plasmids were amplified and isolated as previously described (Shakin-Eshleman et al., 1992).

In Vitro Transcription/Translation/Glycosylation System. Plasmids were linearized with *Xba*I (position 2860) prior to in vitro transcription with T7 RNA polymerase. In vitro transcription and in vitro translation with rabbit reticulocyte lysate and [^{35}S]methionine were performed exactly as described (Shakin-Eshleman et al., 1992). Where indicated, reactions were supplemented with 4 equiv (2 μL) of dog pancreatic microsomes (PM) (Promega Corp., Madison, WI). Three to five microliters of each 15-μL translation reaction were analyzed directly by SDS–PAGE and autoradiography as described (Shakin-Eshleman et al., 1992). Apparent molecular masses of radiolabeled proteins were determined using prestained markers (Bethesda Research Laboratories) including bovine serum albumin (68 kDa), ovalbumin (43 kDa), and carbonic anhydrase (29 kDa).

Quantitation of Core Glycosylation Efficiency. The core glycosylation efficiency of each glycosylation sequon in each RGP protein was defined as the percentage of the protein core glycosylated at that position in the presence of PM. The percentage of each RGP protein which was core glycosylated was determined by densitometric scanning of autoradiographs using Image-Quant version 3.22 (Molecular Dynamics, Sunnyvale, CA). For this quantitation, autoradiographs were selected so that the bands analyzed were in the linear range. The core glycosylation efficiency of each glycosylation sequon in each RGP protein was analyzed in at least two independent experiments.

RESULTS

Construction of RGP Termination Mutants. Wild-type RGP of the ERA strain [RGP(WT)] is a 505 amino acid type I membrane glycoprotein containing a 22 amino acid transmembrane domain, a 44 amino acid cytoplasmic domain, and three Asn-X-Ser/Thr sequons within the extracellular domain, at Asn37 (position 1), Asn247 (position 2), and Asn319 (position 3) (Wunner et al., 1988) (Figure 1A).

The construction of plasmids encoding glycosylation mutants of pRGP(WT), pRGP(---), pRGP(1--), pRGP(-2-), and pRGP(--3), was previously described (Shakin-Eshleman et al., 1992). In each case, the numbers in parentheses indicate the presence of a glycosylation sequon at position 1, 2, or 3,

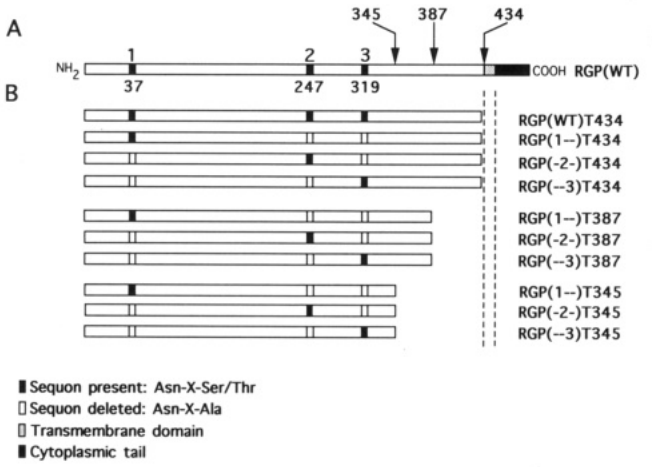

FIGURE 1: Structure of RGP and RGP termination mutants. (A) Structure of RGP(WT). The Asn-X-Ser/Thr sequons at positions 1 (Asn37), 2 (Asn247), and 3 (Asn319) are indicated. The transmembrane domain, which extends from amino acid 440 to amino acid 461, and the C-terminal cytoplasmic tail, which extends from amino acid 462 to amino acid 505, are shown. The positions of amino acids 345, 387, and 434, which represent the sites of termination codons in the RGP termination mutants T345, T387, and T434, respectively, are indicated with arrows. (B) Structures of the RGP termination mutants. The structure of each of the RGP termination mutants is shown. The relative position of the transmembrane domain of RGP, deleted in these mutants, is shown by the vertical dashed lines. Asn-X-Ser/Thr glycosylation sequons are indicated by dark bars; the positions of the corresponding sequons deleted by site-directed mutagenesis are indicated by open bars. The name of each RGP protein is shown at the right. The C-terminal extent of each protein is indicated in the name by the letter, T, followed by the number indicating the amino acid position of the termination codon; the numbers in parentheses indicate the presence of glycosylation sequons at position 1, 2, or 3, while the symbol (-) indicates deletion of the corresponding sequon by site-directed mutagenesis.

while the symbol (-) indicates the deletion of the corresponding sequon by site-directed mutagenesis.

Termination mutants encoding truncated forms of RGP were generated from the plasmids described by introducing termination codons at amino acid positions 345, 387, and 434 (Figure 1B). Termination mutants derived from pRGP(1--), pRGP(-2-), and pRGP(--3) each contain only a single glycosylation sequon, while the termination mutant RGP-(WT)T434 contains all three glycosylation sequons. All of the termination mutants lack a transmembrane domain or membrane anchor; therefore, all of these mutants are fully translocated and released into the lumen of microsomes during their synthesis.

Core Glycosylation of Full-Length RGP Mutants in a Cell-Free System. The expression and core glycosylation of RGP-(WT) and RGP glycosylation mutants were examined using a cell-free transcription/translation/glycosylation system as previously described (Shakin-Eshleman et al., 1992). In this system, RNA transcripts were translated in a rabbit reticulocyte lysate in vitro translation system containing [^{35}S]-methionine. Core glycosylation was accomplished by adding dog pancreatic microsomes (PM) to the translation reaction. The microsomes support amino-terminal signal sequence cleavage and core glycosylation at N-linked glycosylation sequons (Walter & Blobel, 1983). Each mutant described below was assayed in at least two independent experiments with quantitatively similar results.

The pattern of core glycosylation of RGP in the cell-free system was previously shown to be similar to that observed when RGP proteins were expressed in intact cells (Shakin-Eshleman et al., 1992). The cell-free system offers several

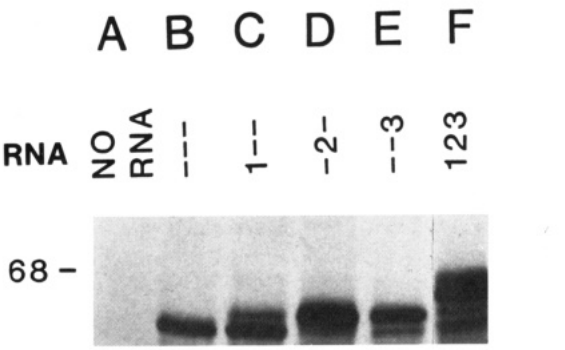

FIGURE 2: Core glycosylation of full-length RGP(WT) and full-length RGP glycosylation mutants. RNA encoding either RGP-(WT) (lane F) or one of the full-length RGP glycosylation mutants, RGP(- - -), RGP(1- -), RGP(-2-), or RGP(- -3) (lanes B–E, respectively), was prepared by in vitro transcription and translated in a rabbit reticulocyte lysate system containing [^{35}S]methionine in the presence of dog pancreatic microsomes (PM). The glycosylation sequons present in each RGP mutant are indicated above each lane. A translation reaction including PM but containing no exogenous RNA is also shown (lane A). Translation products were analyzed directly by SDS–PAGE and autoradiography. The positions of molecular weight (M_r) standards are indicated.

advantages including simplicity, rapidity of analysis, and ease of manipulation. Moreover, since glycosylation occurs in closed microsomal vesicles derived from the RER, the oligosaccharide units on the core-glycosylated proteins are not processed to more heterogeneous forms, simplifying analysis. Other potential effects of the structural alterations of RGP which might complicate an analysis of core glycosylation in intact cells, such as altered protein stability or the kinetics of intracellular transport, do not complicate the analysis in the cell-free system. In addition, since RGP is the major radiolabeled protein produced in the cell-free system, electrophoresis can be performed without immunoprecipitation; this is advantageous because structurally altered forms of RGP may differ in their ability to bind to available antibodies.

Previous studies of RGP in the cell-free system demonstrated that the sequons at positions 2 and 3 are efficiently glycosylated, while the sequon at position 1 is glycosylated at a lower level (Shakin-Eshleman et al., 1992). A similar analysis using full-length RGP(WT) and full-length RGP glycosylation mutants is shown in Figure 2 to permit a direct comparison with the results obtained with termination mutants. In each case, RNA encoding RGP(WT) or an RGP mutant was translated in vitro in the presence of PM. The amount of PM added to the translation reactions was optimized so that >90% of the RGP synthesized was translocated into microsomes, confirmed by signal sequence cleavage, core glycosylation, and protease protection assays (data not shown). Expression of RGP(---), which contains no sequons for N-linked glycosylation, produced a single major radiolabeled species of approximately 56 kDa (Figure 2, lane B). This species represents full-length unglycosylated RGP following removal of the amino-terminal signal sequence. A control translation reaction containing no exogenous RNA (Figure 2, lane A) is shown for comparison. Expression of RGP(1--) (Figure 2, lane C) produced two closely migrating species. The lower band, which comigrated with the product of RGP(---) (Figure 2, lane B), represents full-length, unglycosylated RGP following signal sequence cleavage; the upper band represents the same protein following core glycosylation at position 1. Approximately 35% of RGP-(1--) is core-glycosylated (see Table I). In contrast, expression of either RGP(-2-) or RGP(--3) (Figure 2, lanes D and E, respectively) produced a major radiolabeled species which

Table I: Core Glycosylation Efficiencies (%) at Sequons 1, 2, and 3 in Full-Length RGP and RGP Termination Mutants

position of stop codon	sequon 1 (Asn37)	sequon 2 (Asn247)	sequon 3 (Asn319)
505 (FL)	34, 36	95, 98	94, 94
434	29, 33	99, 99	96, 100
387	32, 36	95, 96	57, 59
345	34, 37	94, 97	66, 68

[a] Full-length RGP (FL) and RGP termination mutants each containing a single sequon at position 1, 2, or 3 were expressed in the cell-free system in the presence of PM. The core glycosylation efficiency of each protein (i.e., the percentage of each protein which was core glycosylated) was quantitated by densitometric scanning of autoradiographs (see Experimental Procedures). The core glycosylation efficiency of each sequon in each protein was determined in two independent experiments; results from both experiments are shown.

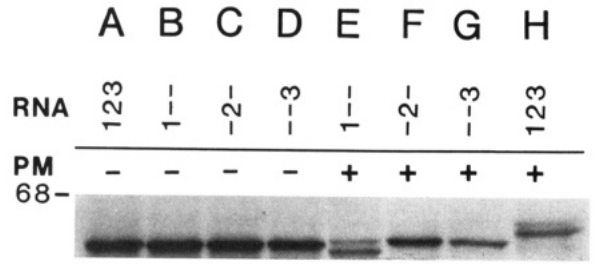

FIGURE 3: Core glycosylation of RGP glycosylation mutants lacking the transmembrane and cytoplasmic domains. RNA encoding RGP-(WT)T434 (lanes A and H), RGP(1- -)T434 (lanes B and E), RGP-(-2-)T434 (lanes C and F), and RGP(- -3)T434 (lanes D and G) was prepared by in vitro transcription and translated in a rabbit reticulocyte lysate system containing [^{35}S]methionine in the presence (lanes E–H) or absence (lanes A–D) of dog pancreatic microsomes (PM). The glycosylation sequons present in each RGP protein are indicated above each lane. Translation products were analyzed directly by SDS–PAGE and autoradiography. The positions of molecular weight (M_r) standards are indicated.

comigrated with the glycosylated product of RGP(1--) and represents full-length RGP following signal sequence cleavage and efficient core glycosylation at position 2 or 3, respectively (see Table I). The autoradiograph shown in Figure 2 was overexposed to permit visualization of the minor unglycosylated proteins in lanes D and E. The expression of RGP(WT) (Figure 2, lane F) produced two major radiolabeled species; these species represent full-length RGP following signal sequence cleavage with the addition of two or three core oligosaccharides (lower and upper bands, respectively) (Shakin-Eshleman et al., 1992).

Core Glycosylation of RGP Mutants Lacking the Transmembrane and Cytoplasmic Domains. To assess the influence of the transmembrane and cytoplasmic domains on the core glycosylation of sequons in RGP, termination mutants were generated from pRGP(1--), pRGP(-2-), pRGP(--3), and pRGP(WT) by the insertion of a termination codon at amino acid position 434 (Figure 1). This termination codon is located 6 amino acids N-terminal to the transmembrane domain. An analysis of the core glycosylation of these mutants is shown in Figure 3. Translation of each of these mutants in the absence of PM (Figure 3, lanes A–D) produced a single major radiolabeled product apparently at 50 kDa; this represents the unglycosylated form of each mutant prior to signal sequence cleavage. The comigration of this product for all four T434 mutants demonstrates that the amino acid differences at the glycosylation sequons (substitution of Ala for Ser or Thr) did not alter the electrophoretic migration of the proteins produced.

The core glycosylation of the T434 mutants was assessed by expressing each protein in the presence of PM (Figure 3, lanes E–H). RGP(1--)T434 (Figure 3, lane E) produced two major radiolabeled species. The lower band, which migrated slightly faster than the proteins produced in the absence of PM (Figure 3, lanes A–D), represents the unglycosylated 433 amino acid translation product following signal sequence cleavage; the upper band, which has an apparent molecular size 3–4 kDa higher, represents the same protein following core glycosylation at position 1. This demonstrates inefficient core glycosylation at position 1 in RGP(1--)T434. Approximately 31% of RGP(1--)T434 was core glycosylated, which is similar to the level of core glycosylation of the full-length RGP(1--) (Figure 2, lane C) (see Table I). In contrast, expression of either RGP(-2-)T434 or RGP(--3)T434 (Figure 3, lanes F and G) in the presence of PM produced a single major radiolabeled product which comigrated with the glycosylated product in lane E. This demonstrates efficient core glycosylation at positions 2 and 3, respectively. This is similar to that seen in the full-length proteins RGP(-2-) and RGP(--3) (Figure 2, lanes D and E, respectively) (see Table I). Translation of RGP(WT)T434 (Figure 3, lane H), which contains all three sequons for N-linked glycosylation, produced two major radiolabeled species, consistent with the addition of two or three core oligosaccharides (lower and upper bands, respectively). This pattern is similar to that seen with full-length RGP(WT) (Figure 2, lane F), consistent with efficient core glycosylation of two of the three sequons in the protein and inefficient glycosylation of the remaining sequon. The results described above reveal that, when compared to the full-length RGP mutants, the deletion of both the transmembrane and cytoplasmic domains had no major effect on the efficiency of core glycosylation at any sequon in RGP.

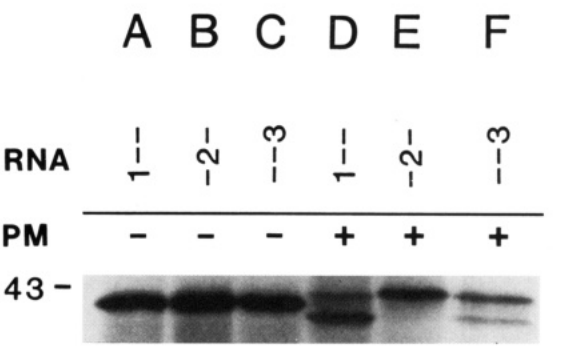

FIGURE 4: Core glycosylation of RGP glycosylation mutants with termination codons at position 345. RNA encoding RGP mutants with termination codons at amino acid position 345 were translated in a rabbit reticulocyte lysate system containing [^{35}S]methionine in the presence (lanes D–F) or absence (lanes A–C) of dog pancreatic microsomes (PM). The glycosylation sequons present in each RGP protein are indicated above each lane. Translation products were analyzed directly by SDS–PAGE and autoradiography. The positions of molecular weight (M_r) standards are indicated.

Core Glycosylation of RGP Mutants with C-Terminal Truncations Extending into the Extracellular Domain. Additional termination mutants were constructed by the introduction of termination codons within the extracellular domain at amino acid positions 345 and 387 (Figure 1). The encoded proteins lack the cytoplasmic and transmembrane domains and portions of the extracellular domain. Analysis of the core glycosylation of the T345 and T387 mutants is shown in Figures 4 and 5, respectively.

Expression of RGP(1--)T345, RGP(-2-)T345, and RGP-(--3)T345 in the absence of PM (Figure 4, lanes A–C) produced a single major radiolabeled translation product approximately 40 kDa in size, representing the unglycosylated translation product prior to signal sequence cleavage. The

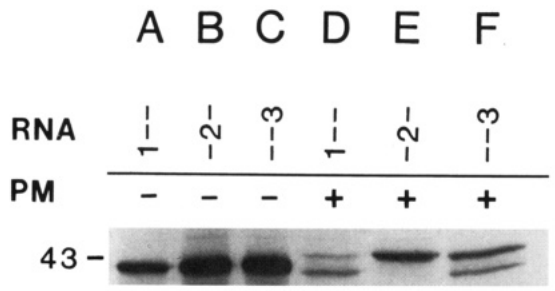

FIGURE 5: Core glycosylation of RGP glycosylation mutants with termination codons at position 387. RNA encoding mutants with termination codons at amino acid position 387 were translated in a rabbit reticulocyte lysate system containing [^{35}S]methionine in the presence (lanes D–F) or absence (lanes A–C) of dog pancreas microsomes (PM). The glycosylation sequons present in each RGP protein are indicated above each lane. Translation products were analyzed directly by SDS–PAGE and autoradiography. The positions of molecular weight (M_r) standards are indicated.

comigration of this product for all three T345 RGP mutants demonstrates that mutations introduced at the glycosylation sequons did not alter the electrophoretic migration of these proteins.

Expression of RGP(1--)T345 in the presence of PM (Figure 4, lane D) produced two major radiolabeled species. The lower band migrated slightly faster than the proteins produced in the absence of PM (Figure 4, lane A) and corresponds to the unglycosylated 344 amino acid translation product following signal sequence cleavage; the upper band, which has an apparent size 3–4 kDa higher, corresponds to the same protein following core glycosylation at position 1. This demonstrates inefficient core glycosylation at position 1 in RGP(--)T345. Approximately 36% of RGP(1--)T345 was core-glycosylated, which is similar to the level of core glycosylation of full-length RGP(1--) (Figure 2, lane C) and the termination mutant RGP(1--)T434 (Figure 3, lane E) (see Table I). In contrast, expression of RGP(-2-)T345 (Figure 4, lane E) in the presence of PM produced a single major radiolabeled product which comigrated with the upper band in lane D, corresponding to the 344 amino acid translation product following core glycosylation at position 2. The efficient core glycosylation at position 2 is similar to that found with full-length RGP(-2-) (Figure 2, lane D) and the termination mutant RGP(-2-)T434 (Figure 3, lane F) (see Table I).

While the efficiency of core glycosylation at positions 1 and 2 was not altered by deletion of amino acids 345–505 (as compared to proteins with the deletion of amino acids 434–505), the efficiency of core glycosylation at position 3 was affected by this additional deletion of 89 amino acids from the extracellular domain. Expression of RGP(--3)T345 in the presence of PM (Figure 4, lane F) produced two major radiolabeled species which comigrated with the species produced from the translation of (1--)T345 in the presence of PM (Figure 4, lane D). As described above, the lower band corresponds to the unglycosylated 344 amino acid translation product following signal sequence cleavage, while the upper band corresponds to the same protein following core glycosylation. These results demonstrate inefficient core glycosylation at position 3 in RGP(--3)T345; in this mutant, approximately 67% of the RGP protein is glycosylated at position 3 (Figure 4, lane F). The lower level of core glycosylation of this termination mutant is in marked contrast to the fully efficient core glycosylation at position 3 in both the full-length RGP(--3) (Figure 2, lane E) and the termination mutant RGP(--3)T434 (Figure 3, lane G) (see Table I). This shows that the efficiency of core glycosylation at a specific sequon is influenced by protein structures more than 26 amino acids C-terminal to the glycosylation sequon.

To further define the influence of C-terminal regions of RGP on core glycosylation, a similar analysis was performed with RGP mutants containing termination codons at amino acid position 387 (Figure 5). Expression of RGP(1--)T387, RGP(-2-)T387, and RGP(--3)T387 in the absence of PM (Figure 5, lanes A–C) produced a single major radiolabeled translation product of approximately 44 kDa, representing the unglycosylated form of each mutant prior to signal sequence cleavage. Expression of the T387 mutants in the presence of PM (Figure 5, lanes D–F) yielded results similar to those seen with the T345 termination mutants (Figure 4). For RGP(--3)T387 (Figure 5, lane F), approximately 58% of the RGP protein was glycosylated at position 3. The inefficient core glycosylation at position 3 of RGP(--3)T387 and RGP(--3)-T345 (Figures 4 and 5, lanes F) contrasts with the efficient glycosylation at position 3 in either full-length RGP(--3) (Figure 2, lane E) or the termination mutant RGP(--3)T434 (Figure 3, lane G) (see Table I). Since RGP(--3)T345, RGP(--3)T387, and RGP(--3)T434 all lack the transmembrane and cytoplasmic domains of RGP, the difference in the efficiency of core glycosylation at the third sequon in these proteins does not reflect a difference in their membrane association. The difference between the efficiency of core glycosylation of RGP(--3)T387 (Figure 5, lane F) and RGP(--3)T434 (Figure 3, lane G) demonstrates that the additional deletion of 47 amino acids from the C-terminus of the extracellular domain (amino acids 387–434) can influence the efficiency of core glycosylation at position 3 (Asn319). This further reveals that the efficiency of core glycosylation at a specific sequon can be influenced by protein structures more than 68 amino acids C-terminal to the sequon.

DISCUSSION

In this report, we test whether the core glycosylation efficiency at specific sequons in RGP is influenced by regions of the protein C-terminal to the sequons. We find that core glycosylation of RGP is not affected by deletion of the C-terminal transmembrane and cytoplasmic domains. In contrast, more extensive C-terminal deletions which truncate the protein by extending into the extracellular domain to amino acid 387 result in a significant decrease in the efficiency of core glycosylation of Asn319. This effect is site-specific since no alterations in core glycosylation are observed at Asn37 or Asn247. This finding, that the presence or absence of regions of RGP more than 68 amino acids C-terminal to Asn319 can influence core glycosylation efficiency at that sequon, would not be predicted by the currently accepted model of core glycosylation; alternative new models which could explain this finding are discussed below.

Most available data support a model in which core glycosylation occurs cotranslationally immediately following the translocation of a sequon into the RER lumen (Figure 6A). This model is supported by finding that the protein translocation channel, the dolichol oligosaccharide donor, and the oligosaccharyltransferase all reside within the RER membrane (Kaplan et al., 1987). Moreover, the active site of oligosaccharyltransferase which recognizes the tripeptide sequon in the translocating nascent chain appears to be either located at the lumenal surface of the RER membrane, or buried within the membrane (Welply et al., 1986). Further support for this model comes from studies which suggest that a sequon has the potential for core glycosylation only after it reaches the lumenal surface of the RER membrane (Glabe

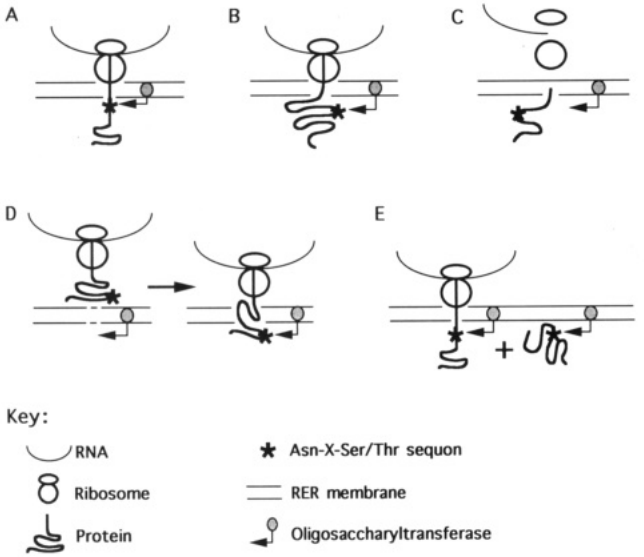

FIGURE 6: Models of N-linked core glycosylation. The schematic diagrams (A–E) illustrate the translation of a protein from mRNA by a ribosome and translocation of the protein across the RER membrane into the RER lumen. Transfer of a core oligosaccharide to an Asn-X-Ser/Thr sequon in the protein by the enzyme oligosaccharyltransferase occurs on the lumenal side of the RER membrane. Model A: Core glycosylation occurs immediately following translocation of a sequon into the RER lumen; according to this model, C-terminal regions of the protein would not influence core glycosylation efficiency. Models B–E illustrate alternative mechanisms by which C-terminal regions of a protein could influence core glycosylation efficiency (see Discussion). Model B: Core glycosylation is delayed until a significant length of the protein is translocated into the RER lumen. Model C: Premature translational termination and release of the protein by the ribosome alter protein folding and the dynamics of protein translocation. Model D: Translocation occurs late in translation or posttranslationally. Model E: Core glycosylation occurs both during and after translocation of the protein into the RER.

et al., 1980; Bergman & Kuehl, 1977). In this model, core glycosylation may or may not be influenced by protein folding within the RER lumen (Bulleid et al., 1992). However, in either case, only those regions of the protein in the immediate vicinity of the sequon or N-terminal to the sequon would be present in the RER lumen when core glycosylation occurs and would be able to influence core glycosylation efficiency. In contrast, regions of the protein significantly C-terminal to the sequon would not be present in the RER lumen when core glycosylation occurs and would not be expected to influence core glycosylation efficiency. According to this model, only approximately 60 amino acids C-terminal to a sequon would have been synthesized (translated) when core glycosylation occurs; of these, approximately 20 amino acids would be located within the RER membrane, and approximately 30–40 amino acids would be located within the ribosome (Glabe et al., 1980; Malkin & Rich, 1967; Sabatini & Blobel, 1970). Regions more than 60 amino acids C-terminal to the sequon, such as those examined in the present report, would not yet have been translated from the mRNA. Therefore, the finding of the present report, that regions of RGP more than 68 amino acids C-terminal to Asn319 influence its core glycosylation efficiency, cannot be readily explained by this model. Four new alternative models are described below which propose different mechanisms by which regions of RGP C-terminal to Asn319 might influence its core glycosylation efficiency (Figure 6B–E).

In the first of these models (Figure 6B), core glycosylation of a sequon might not occur immediately upon its introduction into the RER lumen, but instead might be delayed until additional portions of the protein C-terminal to the sequon are translocated. In this model, since more extensive C-terminal regions of the protein would be present within the RER lumen when core glycosylation occurs, these regions could also influence protein folding and core glycosylation efficiency.

In the second model (Figure 6C), regions of a protein C-terminal to a sequon might influence core glycosylation by influencing the dynamics of protein translocation. For example, the proximity of a sequon to the C-terminus might influence its rate of passage through the RER membrane and into the lumen following translational termination and release of the protein by the ribosome. Premature translational termination, which occurs during translation of the RGP truncation mutants in the present study, might alter core glycosylation efficiency by influencing the time available for protein folding in the vicinity of a sequon or the time during which the sequon remains in the proximity of the glycosylation apparatus. While surveys reveal that many proteins can be glycosylated at sequons very close to their C-termini (Pollack & Atkinson, 1983), it is not known whether the proximity of these sequons to the C-terminus influences their core glycosylation efficiency.

In the third model (Figure 6D), C-terminal regions of a protein could influence core glycosylation of upstream sequons if protein translocation into the RER lumen occurred late in translation (Perara & Lingappa, 1985) or after completion of the protein chain in the cytoplasm (i.e., posttranslationally) (Mueckler & Lodish, 1986; Perara et al., 1986; Rothblatt & Meyer, 1986; Hansen et al., 1986; Waters & Blobel, 1986). In this case, folding of the protein in the cytoplasm might occur prior to its translocation and perhaps be required to trigger its translocation into the RER (Wickner, 1979; Engelman & Steitz, 1981). If a significant length of a protein C-terminal to a glycosylation sequon were synthesized prior to translocation, this C-terminal region might influence protein configuration in the vicinity of the sequon during translocation. Therefore, C-terminal regions could influence the accessibility of the sequon to the glycosylation apparatus and its efficiency of core glycosylation.

In the fourth model (Figure 6E), core glycosylation at a sequon might occur not only during translocation of the nascent protein chain, but also after delivery of the protein into the RER lumen (posttranslationally). The possibility of posttranslational glycosylation was suggested previously by Guan et al. (1985), who observed that core glycosylation efficiency at each of two sequons introduced into rat growth hormone was increased when a membrane anchor was fused to the normally secreted growth hormone protein. These authors postulated that more efficient glycosylation of the membrane-anchored molecules might result from prolonged exposure of the protein to the membrane-bound oligosaccharyltransferase after translation. In the current study, removal of only the transmembrane domain and cytoplasmic tail of RGP did not influence the glycosylation efficiency at any of the three sequons, while more extensive C-terminal truncations decreased the glycosylation efficiency at sequon 3. In this case, the more extensive truncations of RGP could influence posttranslational glycosylation at sequon 3 by altering the folding of the protein and the accessibility of sequon 3 to the glycosylation apparatus after translocation of the protein into the RER lumen.

The models described above (Figures 6B–E) suggest several new mechanisms by which C-terminal regions of RGP could influence core glycosylation efficiency at upstream sequons. Experiments are currently underway with RGP to characterize

the mechanism by which the core glycosylation efficiency of Asn319 is lowered when C-terminal regions of RGP are deleted. Additional studies of other glycoproteins are needed to assess whether C-terminal regions of glycoproteins other than RGP can also influence core glycosylation efficiency.

In these studies, removal of the transmembrane and cytoplasmic domains alone did not affect core glycosylation; this may reflect the distance of these domains from the glycosylation sequons (at least 115 amino acids). Alternatively, amino acids in these domains may not normally influence core glycosylation efficiency of RGP sequons, because they may not impact on protein folding of the extracellular domain. While this deletion (amino acids 434–505) does not influence RGP core glycosylation, it may have other effects on RGP structure and function. Similar deletions in other proteins have pleiotropic effects on protein expression, including alterations in intracellular transport (Raviprakash et al., 1990), stability (Doyle et al., 1986), and oligomerization (Singh et al., 1990; Godet et al., 1991). Moreover, while this deletion did not affect RGP core glycosylation, analogous C-terminal deletions in other proteins influence subsequent processing of core oligosaccharides from high-mannose to complex forms (Finley et al., 1990; Godet et al., 1991; Sveda et al., 1982; Solomon et al., 1991). Further studies of the RGP mutants in intact cells will be needed to further define the impact of these deletions in RGP expression.

The mutants described in this report will be applicable to studies of the biology of rabies virus. RGP serves several important functions in rabies virus infection (Wunner et al., 1988). It is also the major immunologic stimulus of virus neutralizing antibody production in infection (Wunner et al., 1983) and is effective as a vaccine, provided that the protein is glycosylated (Kieny et al., 1984; Wiktor et al., 1984; Yelverton et al., 1983). The termination mutants described here will be helpful in mapping functionally important regions in this protein and in further defining the fine specificity of the immune response in rabies virus infection. In addition, the RGPT434 termination mutants may generate secreted forms of RGP which could facilitate crystallographic analysis of the extracellular, antigenic domain. Of note, a soluble form of RGP, G_s, is produced in infected cells by proteolytic cleavage of RGP after membrane insertion; this truncated protein, which contains the entire extracellular domain of RGP but lacks the cytoplasmic tail and most of the transmembrane domain, retains full antigenicity (Dietzschold et al., 1983; Wunner et al., 1983).

Finally, the finding that C-terminal regions of a protein can influence glycosylation at upstream sequons has practical implications. Advances in molecular biology now make it possible to clone and express a wide variety of genes for research and clinical applications. However, many biologically important proteins require posttranslational modification by processes such as N-linked glycosylation for appropriate expression or function. Therefore, production of novel proteins by genetic engineering requires the ability not only to control the structure of the peptide chain, but also to direct modifications such as glycosylation to appropriate sites within the protein. While the Asn-X-Ser/Thr sequon is required for N-linked glycosylation, little is known about the protein signals which regulate the core glycosylation efficiency at specific sites. Our finding, that regions of a protein significantly C-terminal to a sequon can influence its core glycosylation efficiency, should be considered in the design of chimeric or novel genetically engineered glycoproteins.

ACKNOWLEDGMENT

The authors thank Dr. Phillips Robbins for suggesting the model shown in Figure 6E and Dr. Alan T. Remaley for construction of pRGP(WT)T434.

REFERENCES

Bergman, L. W., & Kuehl, W. M. (1977) *Biochemistry 16*, 4490–4497.
Bulleid, N. J., Bassel-Duby, R. S., Freedman, R. B., Sambrook, J. F., & Gething, M.-J. H. (1992) *Biochem. J. 286*, 275–280.
Dietzschold, B., Wiktor, T. J., Wunner, W. H., & Varrichio, A. (1983) *Viology 124*, 330–337.
Doyle, C., Sambrook, J., & Gething, M.-J. (1986) *J. Cell Biol. 103*, 1193–1204.
Dube, S., Fisher, J. W., & Powell, J. S. (1988) *J. Biol. Chem. 263*, 17516–17521.
Engelman, D. M., & Steitz, T. A. (1981) *Cell 23*, 411–422.
Finley, E. M., Rebbe, N. F., & Hickman, S. (1990) *Arch. Biochem. Biophys. 279*, 395–401.
Glabe, C. G., Hanover, J. A., & Lennarz, W. J. (1980) *J. Biol. Chem. 255*, 9236–9242.
Godet, M., Rasschaert, D., & Laude, H. (1991) *Virology 185*, 732–740.
Green, S., Issemann, I., & Sheer, E. (1988) *Nucleic Acids Res. 16*, 369.
Guan, J.-L., Machamer, C. E., & Rose, J. K. (1985) *Cell 42*, 489–496.
Hansen, W., Garcia, P. D., & Walter, P. (1986) *Cell 43*, 397–406.
Hart, G. W., Brew, K., Grant, G. A., Bradshaw, R. A., & Lennarz, W. J. (1979) *J. Biol. Chem. 254*, 9747–9753.
Hobman, T. C., Qiu, Z., Chaye, H., & Gillam, S. (1991) *Virology 181*, 768–772.
Kaplan, H. A., Welply, J. K., & Lennarz, W. J. (1987) *Biochim. Biophys. Acta 906*, 161–173.
Kelleher, D. J., Kreibich, G., & Gilmore, R. (1992) *Cell 69*, 55–65.
Kieny, M. P., Lathe, R., Drillien, R., Spehner, D., Skory, S., Schmitt, D., Wiktor, T., Koprowski, H., & Lecocq, J. P. (1984) *Nature 312*, 163–166.
Klenk, H.-D. (1990) in *Immunochemistry of Viruses, II* (Van Regenmortel, M. H. V., & Neurath, A. R., Eds.) pp 25–37, Elsevier, Amsterdam.
Kornfeld, R., & Kornfeld, S. (1985) *Annu. Rev. Biochem. 54*, 631–664.
Kronquist, K. E., & Lennarz, W. J. (1978) *J. Supramol. Struct. 8*, 51–65.
Machamer, C. E., & Rose, J. K. (1988) *J. Biol. Chem. 263*, 5948–5954.
Malkin, L. I., & Rich, A. (1967) *J. Mol. Biol. 26*, 329–346.
Marshall, R. D. (1974) *Biochem. Soc. Symp. 40*, 17–26.
Matzuk, M. M., & Boime, I. (1988) *J. Cell. Biol. 106*, 1049–1059.
Matzuk, M. M., Keene, J. L., & Boime, I. (1989) *J. Biol. Chem. 264*, 2409–2414.
Mononen, I., & Karjalainen, E. (1984) *Biochim. Biophys. Acta 788*, 364–367.
Mueckler, M., & Lodish, H. F. (1986) *Cell 44*, 629–637.
Nakai, K., & Kanehisa, M. (1988) *J. Biochem. 104*, 693–699.
Ng. D. T. W., Hiebert, S. W., & Lamb, R. A. (1990) *Mol. Cell Biol. 10*, 1989–2001.
Perara, E., & Lingappa, V. R. (1985) *J. Cell. Biol. 101*, 2292–2301.
Perara, E., Rothman, R. E., & Lingappa, V. R. (1986) *Science 232*, 348–352.
Pizer, L. I., Cohen, G. H., & Eisenberg, R. J. (1980) *J. Virol. 34*, 142–153.
Pless, D. D., & Lennarz, W. J. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 134–138.
Plummer, T. H., & Hirs, C. H. W. (1964) *J. Biol. Chem. 239*, 2530–2538.

Pollack, L., & Atkinson, P. H. (1983) *J. Cell Biol. 97*, 293–300.
Raviprakash, K., Rasile, L., Ghosh, K., & Ghosh, H. P. (1990) *J. Biol. Chem. 265*, 1777–1782.
Rothblatt, J. A., & Meyer, D. I. (1986) *EMBO J. 5*, 1031–1036.
Rothman, J. E., & Lodish, H. F. (1977) *Nature 269*, 775–780.
Sabatini, D. D., & Blobel, G. (1970) *J. Cell Biol. 45*, 146–157.
Sambrook, J., Fritsch, E. K., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Vols. 1–3, CSH Press, Cold Spring Harbor, NY.
Semenkovich, C. F., Luo, C.-C., Nakanishi, M. K., Chen, S.-H., Smith, L. C., & Chan, L. (1990) *J. Biol. Chem. 265*, 5429–5433.
Shakin-Eshleman, S. H., Remaley, A. T., Eshleman, J. R., Wunner, W. H., & Spitalnik, S. L. (1992) *J. Biol. Chem. 267*, 10690–10698.
Singh, I., Doms, R. W., Wagner, K. R., & Helenius, A. (1990) *EMBO J. 9*, 631–639.
Solomon, K. A., Robbins, A. K., & Enquist, L. W. (1991) *J. Viol. 65*, 5952–5960.
Struck, D. K., Lennarz, W. J., & Brew, K. (1978) *J. Biol. Chem. 253*, 5786–5794.
Sveda, M. M., Markoff, L. J., & Lai, C.-J. (1982) *Cell 30*, 649–656.
Tao, M.-H., & Morrison, S. L. (1989) *J. Immunol. 143*, 2595–2601.
Walter, P., & Blobel, G. (1983) *Methods Enzymol. 96*, 84–93.
Waters, M. G., & Blobel, G. (1986) *J. Cell Biol. 102*, 1543–1550.
Welply, J. K., Kaplan, H. A., Shenbagamurthi, P., Naider, F., & Lennarz, W. J. (1986) *Arch. Biochem. Biophys. 246*, 808–819.
Wickner, W. (1979) *Annu. Rev. Biochem. 48*, 23–45.
Wiktor, T. J., Macfarlan, R. I., Reagan, K. J., Dietzschold, B., Curtis, P. J., Wunner, W. H., Kieny, M.-P., Lathe, R., Lecocq, J.-P., Mackett, M., Moss, B., & Koprowski, H. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7194–7198.
Wright, K. E., Salvato, M. S., & Buchmeier, M. J. (1989) *Virology 171*, 417–426.
Wunner, W. H., Dietzschold, B., Curtis, P. J., & Wiktor, T. J. (1983) *J. Gen. Virol. 64*, 1649–1656.
Wunner, W. H., Dietzschold, B., Smith, C. L., Lafon, M., & Golub, E. (1985) *Virology 140*, 1–12.
Wunner, W. H., Larson, J. K., Dietzschold, B., & Smith, C. L. (1988) *Rev. Infect. Dis. 10*, Suppl. 4, S771–S784.
Yelverton, E., Norton, S., Obijeski, J. F., & Goeddel, D. V. (1983) *Science 219*, 614–619.