BACKGROUND PAPER

FUNCTIONS OF THE CORONAVIRUS NUCLEOCAPSID PROTEIN

Paul S. Masters and Lawrence S. Sturman

Wadsworth Center for Laboratories and Research
New York State Department of Health
Albany
NY 12201-0509
USA

In 1962 Caspar and Klug conjectured that self-assembly of equivalent or quasi-equivalent protein subunits and viral nucleic acid produces either icosahedral or helical structures according to the biological functions required (1). Some 15 years later Wengler introduced the terms "transcription helices" and "translation helices" to describe the relationship between helical ribonucleoprotein (RNP) structure and genome function for animal viruses containing single-stranded RNA (2). In transcription helices the RNA genome is transcribed into complementary nucleic acid, without permanent disassembly of the RNP. In translation helices the RNA is liberated from the RNP and translated into protein. To account for the fact that no examples of enveloped viruses containing translation helices had been described, Wengler speculated that "the forces exerted on the viral RNP during budding necessitate the design of a helical RNP of such high stability that RNA cannot be released for translation *in vivo*. Therefore, translation helices will not be present in viruses which obtain a viral membrane by budding".

Today we know this prediction to be wrong: at least one family of viruses, coronaviruses, contains translation helices. However, Wengler's analysis still provides a useful framework for considering some aspects of the molecular biology of these interesting viruses. Most centrally, we do not presently understand how the interactions between the nucleocapsid (N) protein and the viral genome allow the coronavirus RNP to stably acquire its envelope during the maturation process but yet permit the RNA to be released for translation during the early stages of infection. In this overview we will attempt to briefly summarize what is known about coronavirus N proteins and what are the more pertinent outstanding questions.

N PROTEIN STRUCTURE

To date, the deduced amino acid sequences of the N proteins of at least six coronaviruses (MHV, BCV, HCV-OC43, TGEV, HCV229E, and IBV) have been reported (3-5). All have comparable general physical properties. They range from 382 - 455 amino acids in length (M_r 43 - 50,000) and have isoelectric points of 10.3 - 10.7, reflecting a preponderance of basic residues (58 - 72 lysine plus arginine) over acidic residues (44 - 53 glutamate plus aspartate). The basic residues have some degree of localization but are not as densely clustered as in certain viral and cellular nucleic acid binding proteins. The coronavirus N proteins have a 7 - 10% serine content, and many of these residues are clustered. In contrast to the overall basic character of the N molecules, their carboxy termini are markedly acidic: the 45 C-terminal amino acids of each have isoelectric points of 4.5 - 5.3. It is interesting that all of the characteristics noted to this point are also shared by the nucleocapsid proteins of influenza viruses.

Coronaviruses and Their Diseases
Edited by D. Cavanagh and T.D.K. Brown
Plenum Press. New York. 1990

Despite their mutual resemblance, the amino acid sequences of the coronavirus N proteins are generally not very similar. The notable exception to this is a region of 64 - 67 residues, falling within the N-terminal one-third of each molecule, which exhibits a high amount of sequence identity among all the N proteins. The meaning of such a striking degree of conservation is unknown, but it implies that an important function may be assigned to this segment of the molecule.

RNA BINDING

The most salient feature of the N proteins is that they bind to RNA. As such, however, they bear no obvious resemblance to any well-characterized cellular RNA binding protein. In particular, none of the coronavirus N sequences contain regions homologous to the consensus RNA binding sequences common to the family of cellular proteins that includes poly(A) binding proteins, nucleolin and many snRNP proteins (6). A better analogy may be found by comparison with viral RNA binding proteins, specifically the N (or NP) proteins of other single-stranded RNA viruses having helical nucleocapsids. These seem to fall into two classes. The N (NP) proteins of the rhabdoviruses and the paramyxoviruses are neutral or even slightly acidic in their amino acid composition. The complexes they form with their RNA genomes are very stable to conditions of high ionic strength and are highly resistant to the action of RNase. In contrast to these are the N (NP) proteins of the coronaviruses and the orthomyxoviruses. In addition to the above-mentioned physical similarities between the coronavirus and influenza virus N (NP) proteins, both of these form RNPs which are more easily disrupted at high salt concentrations and which afford their RNA genomes relatively little protection against RNase.

Robbins *et al.* developed an RNA overlay protein blot assay (ROPBA) which demonstrated *in vitro* the RNA binding properties of the N protein of MHV but was not specific for viral RNA (7). Stohlman *et al.* made the ROPBA specific for labeled viral RNA by competing out the nonspecific binding of labeled cellular RNA with a large excess of unlabelled cellular RNA (8). These experiments were then extended using synthetic RNA substrates to identify a potential nucleation site for the binding of N to nucleotides 56-65 of the MHV genome and subgenomic RNAs. Interestingly, this region falls within a potential RNA hairpin loop. It remains to be seen whether this will be the only high affinity binding site on the viral RNA. Conversely, the domain(s) of the N protein which participate in RNA binding also have not been determined.

N:N INTERACTIONS

There are potentially two forms of interactions between N molecules in the helical nucleocapsid. The first would be between adjacent monomers bound along the length of the RNA strand. There is presently no evidence for such interactions, although any RNA binding model of nucleation followed by cooperative encapsidation seems to require that they occur. However, one possible interpretation of the relative RNase sensitivity of the coronavirus RNP is that there exists little or no contact between adjacent N monomers. The second type of N:N interaction would be between monomers that become adjacent per each helical turn of the RNP. It is difficult to see how the nucleocapsid could have a helical structure without this sort of interaction. In their study of RNA binding, Robbins *et al.* detected a disulfide-linked multimeric form of the MHV N protein, possibly a dimer or trimer (7). This species, then, may represent one of these types of N:N interaction.

N:M INTERACTIONS

The assembly of coronaviruses is thought to be driven by interactions between the N protein in the RNP and the large cytoplasmic domain of the M protein, which may act in a manner analogous to the matrix proteins of rhabdoviruses and paramyxoviruses. Sturman *et al.* found that in NP40-disrupted MHV virions a temperature-dependent association between M and the RNP could be demonstrated *in vitro* (9). However, much work remains to be done to define the regions of the N and M proteins which participate in viral maturation.

PHOSPHORYLATION

The N proteins of coronaviruses are known to be phosphorylated both intracellularly and in assembled virions. Moreover, for MHV a virion-associated protein kinase has been described by Siddell *et al.* which can transfer additional phosphate from ATP to the N protein *in vitro* (10). This enzyme appears to be cyclic AMP-independent and may be of either viral or host origin. The phosphate linkage, *in vitro* and *in vivo*, has been shown to be to serine residues, and this has prompted comment about the high serine content of the coronavirus N proteins. However, an HPLC analysis of the tryptic phosphopeptides of MHV N proteins suggests that, despite the large number of potential target residues, N phosphorylation may occur at only some small number of sites (1 - 3) (11). The exact number and location of the N protein phosphoserines and the identity of the responsible protein kinase(s) are open questions at present.

Perhaps more intriguing is the related question of the function of N protein phosphorylation. Stohlman *et al.* have shown that the MHV N protein is phosphorylated in the cytoplasm of infected cells within 10 to 20 minutes of its synthesis and it concomitantly becomes associated with a membrane fraction of the cells (12). It has been speculated by a number of workers that N phosphorylation may govern the tightness of the association between N and RNA, and it may thus be a regulator of assembly. However, there is as yet no evidence that bears on this possibility.

PARTICIPATION IN RNA SYNTHESIS

Compton *et al.*, working with an *in vitro* RNA synthesizing system prepared from MHV-infected cells, found that RNA synthesis was greater than 90% inhibited by antibodies to N (13). This implies a critical role for N protein in the transcription and replication of viral RNA. N may act in one or both of two capacities: firstly, by encapsidation, as an antiterminator or protector of the nascent RNA strand; secondly, as a component of the template, associating with the proteins that constitute the viral RNA polymerase. Compton *et al.* have shown that the RNA product synthesized in their *in vitro* system is encapsidated by N protein (13). However, the data of Perlman *et al.* indicate that the pools of free N protein in infected cells are quite large, and thus free N may not be rate limiting for MHV RNA synthesis even though it clearly participates in the process (14).

Since most newly synthesized coronavirus RNA is of positive polarity, then N protein must be directly associated with the negative-stranded antigenome if N functions as an essential template component. At present, it has not been established that this is the case. Baric *et al.* have shown, contrary to expectations, that monoclonal antibodies to N coprecipitate <u>all</u> intracellular viral RNAs of MHV, including positive-stranded subgenomic messages, leaders larger than 65 nucleotides and the negative-stranded antigenome (15). In the latter case, however, the negative-stranded RNA may be coprecipitating by virtue of its association with genome in a double-stranded RNA replicative form. In the modified ROPBA of Stohlman *et al.* N protein failed to bind to a 1.1 kb RNA complementary to the 5' end of the MHV genome (8). Thus it is uncertain whether N directly participates as part of the transcription template in a manner similar to that observed with rhabdoviruses, paramyxoviruses and orthomyxoviruses.

OTHER N PROTEIN FUNCTIONS

A few other aspects of coronavirus N protein function warrant mention because they may point to additional potentially exciting areas of research. First, the finding by Baric *et al.* that N is complexed to at least part of all MHV mRNAs (15) raises the possibility that it plays some role in translation. N protein cannot be strictly required for translation since naked coronavirus RNA is infectious. Nevertheless, N may in some way act as a translational enhancer. Second, many coronavirus N genes encode other polypeptides in internal overlapping reading frames. However, it is not known whether any of these proteins are actually expressed *in vivo* and, if so, what may be their functions. Finally, the N protein of MHV has been shown to migrate to the nucleus of infected cells. This translocation is not essential to a productive infection because MHV is capable of replicating in enucleated cells. Nuclear migration may mean, however, that the N protein plays some modulatory role in

host cell gene expression. In conclusion, then, it is clear that the molecular biology of coronavirus N proteins contains numerous areas deserving much further exploration.

REFERENCES

1. D.L.D. Caspar and A. Klug, Physical principles in the construction of regular viruses, Cold Spring Harbor Sympos. Quant. Biol., 27: 1 (1962).
2. G. Wengler, Structure and function of the genome of viruses containing single-stranded RNA as genetic material: the concept of transcription and translation helices and the classification of these viruses into six groups, Current Topics Microbiol. and Immunol., 70: 239 (1977).
3. W. Spaan, D. Cavanagh, and M.C. Horzinek, Coronaviruses: structure and genome expression, J. Gen. Virol., 69: 2939 (1988).
4. S.S. Schreiber, T. Kamahora, and M.M.C. Lai, Sequence analysis of the nucleocapsid protein of human coronavirus 229E, Virology, 169: 142 (1989).
5. T. Kamahora, L.H. Soe, and M.M.C. Lai, Sequence analysis of the nucleocapsid gene and leader RNA of human coronavirus OC43, Virus Res., 12: 1 (1989).
6. C.C. Query, R.C. Bentley, and J.D. Keene, A common RNA recognition motif identified within a defined U1 RNA bindingdomain of the 70K U1 snRNP protein, Cell, 57: 89 (1989).
7. S.G. Robbins, M.F. Frana, J.J. McGowan, J.F. Boyle, and K.V. Holmes, RNA binding proteins of coronavirus MHV: detection of monomeric and multimeric N protein with an RNA overlayprotein blot assay, Virology, 150: 402 (1986).
8. S.A. Stohlman, R.S. Baric, G.W. Nelson, L.H. Soe, L.M. Welter, and R.J. Deans, Specific interaction between coronavirus leader RNA and nucleocapsid protein, J. Virol., 62:4288 (1988).
9. L.S. Sturman, K.V. Holmes, and J. Behnke, Isolation of coronavirus envelope glycoproteins and interaction with the viral nucleocapsid, J. Virol. 33:449 (1980).
10. S.G. Siddell, A. Barthel, and V. Ter Meulen, Coronavirus JHM: a virion-associated protein kinase, J. Gen. Virol., 52: 235 (1981) .
11. S.M. Wilbur, G.W. Nelson, M.M.C. Lai, M. McMillan, and S.A. Stohlman, Phosphorylation of the mouse hepatitis virus nucleocapsid protein, Biochem. Biophys. Res. Commun., 141: 7 (1986).
12. S.A. Stohlman, J.O. Fleming, C.D. Patton, and M.M.C. Lai, Synthesis and subcellular localization of the murine coronavirus nucleocapsid protein, Virology, 130: 527 (1983).
13. S.R. Compton, D.B. Rogers, K.V. Holmes, D. Fertsch, J. Remenick, and J.J. McGowan, In vitro replication of mouse hepatitis virus strain A59, J. Virol., 61: 1814 (1987).
14. S. Perlman, D. Ries, E. Bolger, L.-J. Chang, and C.M. Stoltzfus, MHV nucleocapsid synthesis in the presence of cycloheximide and accumulation of negative strand MHV RNA, Virus Res., 6: 261 (1987).
15. R.S. Baric, G.W. Nelson, J.O.Fleming, R.J. Deans. J.G. Keck, N. Casteel, and S.A. Stohlman, Interactions between coronavirus nucleocapsid protein and viral RNAs: implications for viral transcription, J. Virol., 62: 4280 (1988).