Vol. 154, No. 1, 1988 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 15, 1988 Pages 318-325

NUCLEOTIDE SEQUENCE PREDICTS CIRCULARITY AND SELF-CLEAVAGE OF 300-RIBONUCLEOTIDE SATELLITE OF ARABIS MOSAIC VIRUS

J.M. Kaper, M.E. Tousignant and G. Steger*

Microbiology and Plant Pathology Laboratory, Plant Sciences Institute, ARS, U.S. Department of Agriculture, Beltsville, Maryland 20705

*Institut für Physikalische Biologie, Universität Düsseldorf, FRG

Received June 3, 1988

SUMMARY: The nucleotide sequence of the satellite of arabis mosaic virus was determined using the satellite RNA encapsidated in virions. The 300-nucleotide long sequence showed extensive homology (50%) with that of the 359-nucleotide satellite RNA of tobacco ringspot virus, which occurs both in a linear and a circular form. This homology also revealed the presence of conserved sequences believed to mediate self-cleavage of the latter as well as other viral satellite RNAs. A circular form of the arabis mosaic virus satellite can be isolated from infected tissues and partially converts to the linear form upon elution from denaturing gels. © 1988 Academic Press, Inc.

INTRODUCTION: Viral satellites are small nucleic acids (size range 300-1700nt) that can replicate in certain hosts only in the presence of their "helper virus" nucleic acids with which they have little, if any, sequence homology. By far the great majority of viral satellites are single-stranded RNAs associated with plant viruses (1). Viral satellites are of considerable interest due to their capability to modulate the severity of disease symptoms caused by their helper viruses. This symptom modulation has been observed to range from complete suppression to severe aggravation of symptoms, leading to death of the host in some cases (2,3). While symptom suppression can be rationalized by the satellite's own rapid replication at the expense of the virus (4), severe symptom aggravation must result from some form of expression of the nucleotide sequence of the satellite during the course of its replication. Such considerations have made these molecular level parasitic RNAs premier objects for studies that attempt to link RNA structure to biological function (2,3).

Although all known plant viral satellites exist as linear (L) or circular (C) single-stranded RNAs encapsidated in their own or their helper

Abbreviations: L, linear; C, circular; TNA, total nucleic acid; PAGE, polyacrylamide gel electrophoresis; sTBRV, satellite of tomato blackring virus; sTobRV, satellite of tobacco ringspot virus; sArMV, satellite of arabis mosaic virus; VTMoV, velvet tobacco mottle virus.

0006-291X/88 $1.50
Copyright © 1988 by Academic Press, Inc.
All rights of reproduction in any form reserved.

viral coat proteins, variant structural forms can be found unencapsidated in the infected cell. These often are concatenated L (and probably also C) satellite RNA (+) and (-) strands [(+) polarity is assigned to the encapsidated strand], while in some cases large intracellular accumulation of satellite double-stranded RNA signifies unusually active synthesis of (-) stranded satellite RNA (4).

To date the two best characterized satellites associated with a group of plant viruses designated as Nepoviruses (5), are the satellite of tomato blackring virus (sTBRV), and the satellite of tobacco ringspot virus (sTobRV). Biologically the concurrent replication of sTBRV with its helper virus has shown no disease modulating effects, but with sTobRV dramatic symptom attenuation has been reported with some strains of the helper virus (6). Structurally, sTBRV and sTobRV are also quite different in that the former is encapsidated as a 1375nt L RNA (7), and the latter as a 359nt L RNA (8) which coexists with an intracellular C form (9) and lesser amounts of concatenated satellite strands of both polarities (10).

In 1983, a new nepoviral satellite was isolated from hop with nettlehead disease caused by arabis mosaic virus (ArMV) (11), a virus also important in certain fruit crops (5). In contrast to the two nepoviral satellites above, the satellite of ArMV (sArMV) seems to aggravate viral disease symptoms in Chenopodium quinoa, and its association with ArMV in hop was suspected to be the cause of the nettlehead disease (11). Here we report 1) the nucleotide sequence of sArMV, 2) the coexistence of an encapsidated L with an intracellular C form of the satellite, 3) the relationship of this coexistence to certain conserved sequences shared by sArMV, STobRV and certain other small RNA molecules, and thought to be instrumental in the self-cleavage properties of these RNAs (12).

MATERIALS AND METHODS

Virus source, growth, and purification. ArMV was a gift of Dr. M.F. Clark, East Malling Research Station, U.K. Our specimen was derived from the nettlehead isolate ArMV-Ta from hop, mechanically transmitted and subsequently maintained in Chenopodium quinoa. We have continued propagation of the virus in Chenopodium quinoa in a growth chamber at 16-18°, with a photoperiod of 16 h and a light intensity of 1,000 lux. Virus purification was according to Davies and Clark (11).

Isolation of sArMV from virions or from infected tissues. Virion-encapsidated nucleic acids were extracted with phenol-SDS (13), and were fractionated by rate zonal ultracentrifugation for 3 h at 40,000 rpm (Beckman[2] SW-41 rotor) at 5°, in a 0.2-0.8 M sucrose gradient in TAE buffer (TAE = 40 mM Tris-acetic acid, pH 7.8, 20 mM sodium acetate, 2 mM EDTA) plus 0.1% SDS. Separation of sArMV RNA from viral RNAs was by fractionation of the gradients through the top of the centrifuge tubes using an ISCO density gradient fractionator. To

[2] Mention of a commercial company or specific equipment does not constitute its endorsement by the U.S. Dept. of Agriculture over similar equipment or companies not named.

obtain total nucleic acids (TNA) from ArMV-infected Chenopodium, 5 g of tissue was powdered in a mortar and pestle with liquid nitrogen until the tissue became light green. This tissue was extracted with 12 ml of extraction (E) buffer (0.1 M Glycine-NaOH, pH 9.0, 0.1 M NaCl, 10 mM EDTA, 2% SDS, 1% sodium lauryl sarcosine), 12 ml of water-saturated phenol, containing 0.1% 8-hydroxyquinoline, and 6 ml chloroform. Phases were separated by low speed centrifugation, and the nucleic acids in the aqueous phase were subjected to several cycles of ethanol precipitation and resuspension in E buffer (final volume 1 ml). To separate sArMV from viral and cellular nucleic acids, appropriate aliquots of this TNA extract were subjected to polyacrylamide gel electrophoresis (PAGE) on 6% (39:1 acrylamide:bisacrylamide containing 7 M urea, 90 mM Tris-borate, pH 8.3, and 2.5 mM EDTA) slab gels at 30 V/cm for 1.5 h. Marker-identified sArMV bands C and L, which are clearly distinguishable from plant nucleic acid bands in TNA extracts from healthy Chenopodium (Fig. 1A), were excised, crushed in 0.5 M ammonium acetate, 1 mM EDTA, 0.1% SDS, pH 7, and eluted overnight at 4°. Removal of acrylamide was accomplished by low speed centrifugation and ethanol precipitation of the nucleic acids, or by rate zonal ultracentrifugation on sucrose gradients, where the acrylamide stays on top of the gradient.

Nucleotide sequence determination of sArMV. The 5' terminus of the L form of sArMV from virions or from infected tissue was labeled using [γ-^{32}P]ATP and T4 polynucleotide kinase. Labeling of the 3' terminus of sArMV with [5'-^{32}P]pCp was carried out following maleic acid treatment (8) to hydrolyze a 2':3' cyclic phosphodiester group that blocked accesibility to this end of the molecule. This enabled determination of the nucleotide sequences of the 5' and 3' thirds of the molecule, respectively, including the 3' terminal nucleotide, using procedures previously described (14,15). These sequences were overlapped with those of 5' end-labeled sArMV fragments generated by partial digestion with RNase T1. The bulk of the sArMV sequence was subsequently confirmed using the dideoxy chain termination method as previously described (15), and an oligodeoxynucleotide complementary to the eighteen 3' end terminal nucleotides as the primer.

Nucleotide sequence comparisons and secondary structures. Nucleotide sequences were aligned using a dot-plot matrix filtered by a window of w=11 nt, a stringency of s=8 nt, and a computer program analogous to the one described by Collins and Coulson (16), with the remaining homologous nucleotides aligned manually. The most favorable secondary structures at 25° were calculated using a modified Zuker-Nussinov algorithm as previously described (17).

RESULTS

Fig. 1A represents a PAGE comparison of TNA extracts of tissues of Chenopodium quinoa infected with ArMV (lane 2) and healthy plants (lane 1). It shows the conspicuous presence of two fast-migrating nucleic acids in the infected plant extract, which were designated C and L. Fig. 1B compares an ArMV-infected Chenopodium TNA extract (lane 1) with a mixture of the RNAs extracted from ArMV virions (lane 2), and shows that L comigrates with the fastest RNA in the viral mixture. Davies and Clark (11) previously identified this component as the sArMV. In addition, the mixture contains the two genomic RNAs (upper two bands in lane 2), and a fourth component comigrating with an RNA in the TNA extract that is sequence-related to sArMV (compare Fig. 1C, lanes 2 and 6). Purified sArMV (Fig. 1C, lane 1) served as template to prepare a cDNA probe, using a random priming method for the reverse transcription (18). Fig. 1C shows the use of this probe in Northern hybridization of blots to which purified sArMV (lane 5), an ArMV-infected Chenopodium TNA extract

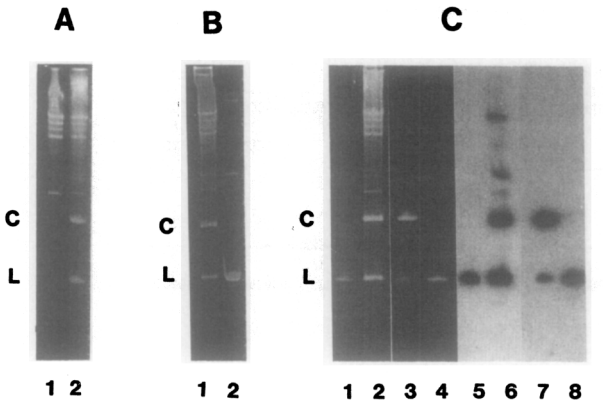
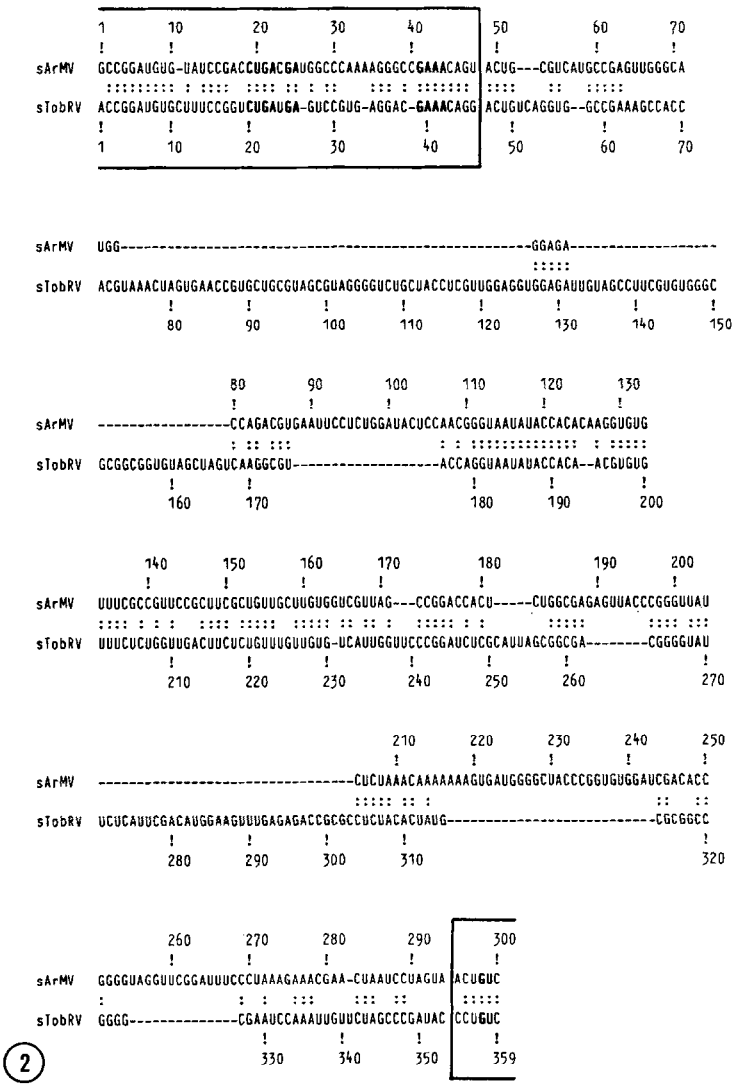


Fig. 1. Comparative 6% PAGE analysis under denaturing conditions of RNA mixtures from ArMV infections in Chenopodium quinoa. A) TNA extracts of uninfected (lane 1) and ArMV-infected (lane 2) plants. B) TNA extract from ArMV-infected plants (lane 1) and purified ArMV virions (lane 2). C) sArMV isolated from virions (lanes 1 and 5); TNA extract of ArMV-infected plants (lanes 2 and 6); C band eluted from a TNA gel (lanes 3 and 7); L band eluted from a TNA gel (lanes 4 and 8). Lanes 1-4 and 5-8 represent the same gel ethidium bromide stained, and hybridized to sArMV cDNA after Northern transfer, respectively.

(lane 6), gel-eluted components C (lane 7) and L (lane 8), had been transferred from 6% gels, following denaturing PAGE and ethidium bromide staining (see lanes 1-4). Clearly, components C and L in the TNA extract have sequence homology with sArMV. In addition, slower migrating sArMV sequence-related components can be observed in this extract (lane 6). Eluted C consistently shows the presence of some L (lanes 3 and 7), even after repeated re-elution (not shown), but L shows only a single band (lanes 4 and 8). C and L appear to represent the circular and linear forms of sArMV (see DISCUSSION).

The nucleotide sequence of sArMV, initially determined using encapsidated sArMV (Fig. 1C, lane 1) and subsequently confirmed with the L component eluted from gels with TNA extract of infected Chenopodium (Fig. 1C, lane 4), is shown in Fig. 2, where it is aligned with the published sequence of sTobRV (8) (Fig. 4 shows the sequence in the form of its most favorable secondary structure, also compared with that of sTobRV). The sequence alignment of the 300 nt sArMV with the 359 nt sTobRV (Fig. 2) reveals that relative to sArMV the sequences are 50% homologous overall (42% relative to sTobRV), whereas only 25% homology can be expected from aligned random sequences of 350 nt. This homology is primarily concentrated in two regions: positions 296-300/1-63 with 70% homology and positions 105-204 with 65% homology (sArMV numbering used throughout, unless mentioned otherwise). Positions 295-300/1-48 (and equivalent sTobRV positions) have been boxed off for both satellites because this region contains the conserved sequences (highlighted in Fig. 2) which, together with three stems and loops, form a consensus configuration known as the "hammerhead" structure. This structure is thought to be instrumental in



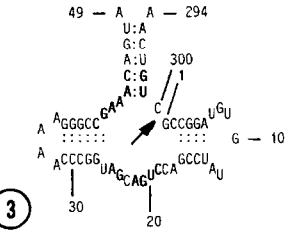


Fig. 2. Nucleotide sequence of sArMV aligned with that of sTobRV. Boxed regions encompass the conserved (bold-faced) and auxiliary sequences that can form the hammerhead structures of the two satellites.

Fig. 3. Hammerhead structure of sArMV with conserved sequences in bold face. Arrow indicates the proposed self-cleavage site of a C form (or oligomer) of sArMV, which coincides with a putative junction of the 3' and 5' terminal nucleotides of the L form.

promoting the self-cleavage properties of certain small RNA molecules (12). The energetically most favorable hammerhead structure that can be drawn for sArMV is shown in Fig. 3.

Secondary structures for the C forms of sArMV and sTobRV are shown in Fig. 4. Their minimum free energies calculated at 25° are -409 and -489 kJ/mol, respectively, and are essentially identical if the different lengths of the two satellites are taken into account ($\Delta G/N$ = -1.36 kJ/mol). The resemblance of the two structures, with their central rod-like, highly base-

paired stems, interrupted by four or five stem-loop structures, can be accounted for mainly by the extended stretches of homology mentioned above between the primary structures of sArMV and sTobRV (Fig. 2). These stretches of homology encompass the middle part of the central stems plus the two relatively uninterrupted stem-loops to the right of this middle part (Fig. 4).

DISCUSSION

In the chronology of this investigation the determination of the nucleotide sequence of encapsidated sArMV, with its homology to the sequence of sTobRV, and with the conserved sequences of the hammerhead structure (Figs. 2 and 3), predicted the existence of a C form of sArMV. As in the case of sTobRV (9), such a C form was detected in ArMV-infected Chenopodium tissues

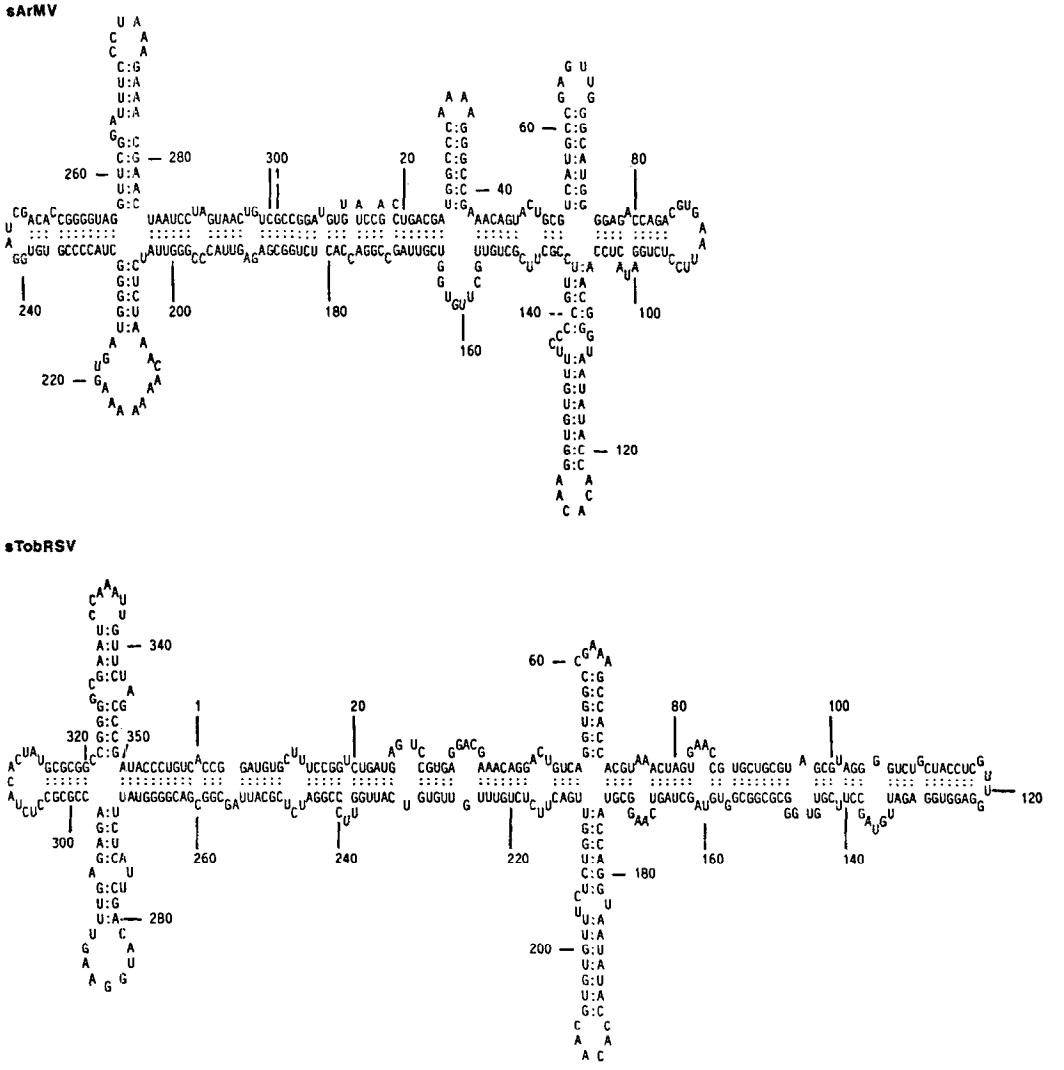


Fig. 4. Most stable secondary structures of sArMV (upper) and sTobRV (lower) at 25°.

(Fig. 1). It also predicted the probable interrelationship of the C and L forms of sArMV via a unique cleavage of the $--C_{+1}\text{-}G_{-1}--$ linkage in a putative hammerhead structure of C sArMV (Fig. 3). Detailed evidence for the validity of these contentions will be presented in a separate publication. C's circularity can be demonstrated via its unique electrophoretic properties in denaturing PAGE, its resistance to 5' end radiolabeling as compared with L, and other properties, while the ability of C to undergo self-cleavage at the predicted junction will be shown with a detailed analysis of a reaction mixture following gel elution of the type shown in Fig. 1, lane 3.

Preliminary comparisons of the sArMV sequence with the published sequences of a number of other viral satellites, which included sTBRV (7), the cucumoviral satellites D-CARNA 5 and PARNA 5, the satellite of tobacco necrosis virus, the viroid-like circular satellite of velvet tobacco mottle virus (also referred to as VTMoV RNA 2 or the "virusoid" of VTMoV), and the satellite-like RNAs 3 and 4 of bean necrotic yellow vein virus (see ref. 3 for quotations of the original publications), revealed only significant homology with VTMoV RNA 2. This homology (69%) was concentrated between positions 288-300/1-52 in sArMV and positions 47-96 in VTMoV RNA 2. Such an homology in the corresponding region had already been reported for sTobRV and VTMoV RNA 2 (8), and clearly relates to a hammerhead structure in the latter molecule, believed to predict its self-cleaving properties (19).

The extensive additional homology between sArMV and sTobRV (Fig. 2), which seems mainly responsible for the striking similarity of their energetically most favorable secondary structures (Fig. 4), may have biological significance. However, precisely in what way is not clear at the present time. ArMV and TobRV represent two different Nepovirus subgroups that have shown no serological relationship to each other (5). Their genomic RNAs are very similar in size, but there is no published information as to their possible sequence homologies. Nevertheless, we have performed some limited tests to check whether the sequence homology between the two satellites might signify that they can also be supported by each other's helper viruses. The result of this test was negative, in that TobRV failed to support sArMV replication in bean while ArMV failed to support sTobRV replication in Chenopodium. However, more unequivocal tests would have to include other host plants and the use of protoplasts. Such experiments are in progress.

ACKNOWLEDGEMENTS: We thank M.T. Steen for assistance with the initial sequence comparisons; L. Tinsley and L. Geletka for technical assistance.

REFERENCES

1. Francki, R.I.B. (1985) Ann. Rev. Microbiol. 39, 151-174.
2. Kaper, J.M., and Tousignant, M.E. (1984) Endeavour New Series 8, 196-200.

3. Kaper, J.M., and Collmer, C.W. (1988) in RNA Genetics (Domingo, E., Holland, J.J., and Ahlquist, P., eds.), Vol. III, pp. 171-194, CRC Press, Boca Raton, Florida.
4. Kaper, J.M. (1984) in Control of Virus Diseases (Kurstak, E., ed.), pp. 317-343, Marcel Dekker, New York.
5. Murant, A.F. (1981) in Handbook Of Plant Virus Infections and Diagnosis (Kurstak, ed.), pp. 198-238, Elsevier/North Holland, Amsterdam, New York.
6. Gerlach, W., Buzayan, J., Schneider, I.R., and Bruening, G. (1986) Virology 151, 172-185.
7. Meyer, M., Hemmer, O., and Fritsch, C. (1984). J. Gen. Virol. 65, 1575-1583.
8. Buzayan, J., Gerlach, W., Bruening, G., Keese, P., and Gould, A.R. (1986) Virology 151, 186-199.
9. Linthorst, H.J.M. and Kaper, J.M. (1984) Virology 137, 206-210.
10. Kiefer, M.C., Daubert, S.D., Schneider, I.R., and Bruening, G. (1982) Virology 121, 262-273.
11. Davies, D.L., and Clark, M.F. (1983) Ann. Appl. Biol. 103, 439-448.
12. Forster, A.C., and Symons, R.H. (1987) Cell 50, 9-16.
13. Kaper, J.M. and West, C.K. (1972) Prep. Biochem. 2, 251-263.
14. Collmer, C.W., Tousignant, M.E., and Kaper, J.M. (1983) Virology 127, 230-234.
15. Kaper, J.M., Tousignant, M.E., and Steen, M.T. (1988) Virology 163, 284-292.
16. Collins, J.F., and Coulson, A.F.W. (1987) in Nucleic Acid and Protein Sequence Analysis (Bishop, M.J., and Rawlings, C.J., eds.), pp. 326-329, IRL Press, Oxford.
17. Steger, G., Hofmann, H., Förtsch, J., Gross, H.J., Randles, J.W., Sänger, H.L., and Riesner, D. (1984) J. Biomolec. Struct. Dyn. 2, 543-571.
18. Taylor, J.M., Illmensee, R., and Summers, J. (1976) Biochim. Biophys. Acta 442, 324-338.
19. Forster, A.C., and Symons, R.H. (1987) Cell 49, 211-220.