Vol. 134, No. 1, 1986 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
January 14, 1986 Pages 372-377

SEQUENCE HOMOLOGY BETWEEN THE COAT PROTEINS OF DNA AND RNA PLANT VIRUSES

J. K. MOHANA RAO

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907

Received November 25, 1985

SUMMARY: The coat proteins of the leafhopper transmitted gemini viruses, viz., maize streak virus and wheat dwarf virus that contain a single circle of DNA in their genomes are shown to bear good amino acid sequence homology with the coat protein of an RNA plant virus, the satellite tobacco necrosis virus. It is suggested that for icosahedral assemblage, the coat protein architecture transcends genomic preferences.

Among viruses, the geminiviruses (for reviews, see refs. 1-3) are uniquely characterized by their occurrence as morphologically distinct geminate particles that consist of two fused incomplete icosahedra and contain 22 pentameric capsomeres of approximately 18 X 30 nm in size. Unlike many other plant viruses, the genomes of these viruses contain single stranded DNA circles. Geminiviruses are transmitted either by whiteflies or by leafhoppers. Whereas the serologically related whitefly transmitted viruses (African cassava latent virus - CLV, tomato golden mosaic virus - TGMV, and bean golden mosaic virus - BGMV) were shown to possess a bipartite genome, the leafhopper transmitted viruses (maize streak virus - MSV and wheat dwarf virus - WDV) contain only a single circle of DNA in its genome. The transcription in the case of MSV seems to be bidirectional. The DNA sequences of all the five viruses mentioned above are available (4-9). Several open reading frames that could code for long polypeptides exist in these DNAs. The probable coat proteins in CLV, TGMV and MSV were identified and sequence

0006-291X/86 $1.50

homology among these were independently established by many groups (9-11).

Details concerning the three-dimensional architecture of many crystalline plant and animal viruses have been published (12-16). The underlying feature of all these structures is the existence of a common structural motif, viz., a cluster of antiparallel beta strands twisted into a right-handed helix, called the "Swiss roll" or the "jelly roll". Since the geminate particles also consist of icosahedra, albeit incomplete, it is expected that the structural fold of the geminivirus coat proteins has the "Swiss roll" beta barrel topology. In the present note, an attempt is made to show that MSV and WDV bear good sequence homology with the RNA plant virus, satellite tobacco necrosis virus (STNV). Since the geminivirus coat proteins themselves bear reasonable sequence homology (11), it may be inferred that all the geminivirus coat proteins probably possess the "Swiss roll" structural motif.

METHODS

In this analysis, searches were made based on the occurrence of identical residues, on the exchange scores of the log-odds matrix of Dayhoff (17), and, on the correlations between the residue physical parameters. When comparing two proteins, all segments of a particular length in the first protein were compared against all segments of the same length in the second protein. In the first case, the number of exactly identical residues in the two segments was counted. In the second instance, where accepted point mutations are to be studied, the scores for the corresponding exchanges between residues of the two candidate segments were read and averaged. Finally, the parametric values for the two segments were found and a correlation coefficient between the two segments was calculated. Similar correlations were evaluated for other parameters also and a mean value was determined so that no one particular parameter dominates thereby prejudicing the analysis. In this analysis, five parameters were used - three that describe the topological traits, viz., the Chou-Fasman alpha helical, beta strand and reverse turn propensity parameters (18-20) and two that describe the polar nature of the amino acids, viz., the residue polarity (21) and a consensus hydrophobicity (22). Experience has taught that for a segment length of 20, significance may be attached to scores of five and above for identical residues, to scores of 10.5 and above with the MDM matrix and to correlation coefficients of 0.4 and above for the physical parameters. The tentatively identified regions of possible significance were

further analysed to establish and maximize the sequence homology. Examples of the successful use of the methods described here are available (11,23). All the calculations were performed on the CYBER 205 and CDC Computers at the Purdue University using programs written by the author.

RESULTS AND DISCUSSION

First of all, the finding of MacDowell et al. (9) that MSV and WDV are highly homologous was confirmed. Among the crystalline plant viruses whose structures were established, the structure of STNV is the most compact one as it has only 195 amino acid residues in its coat protein (24). Southern bean mosaic virus (SBMV) and tomato bushy stunt virus, though bearing good sequence homology between themselves (25), do not bear amino acid sequence homology with STNV in spite of a high degree of structural homology. When the sequence of various geminiviruses was compared with those of the above, the STNV:MSV and the STNV:WDV comparisons seemed most promising and therefore were pursued further. A distinguishing feature of these comparisons was that the homology of one tended to stand out where the other was weak and vice versa thus complementing each other. An alignment of MSV and WDV with STNV was made and a consensus sequence of both MSV and WDV (called MSWDV) that would best align with STNV was obtained. This sequence of MSWDV when compared against that of STNV confirmed that the initial alignment was correct and the homology is almost unambiguous. Occasionally, significant variations in the sequence are observed among different strains of the same virus. Therefore, the choice of a consensus sequence for two different, yet homologous, viruses seems justifiable.

The final alignment of the sequences of STNV, MSV, and, WDV is shown in Fig. 1 and some corroborating statistics are presented in Table 1. In all these comparisons, the average value of the minimum base change per codon (MBC/C) for the

```
          ooooooooooo  bbbbbbbbbbbbb      ccccc                  dddddd
STNV      AVKRMI..NTHLEHKRFALINSGNTNATA..GTVQNLSNGIIQGDDINQRSGDQVRIV
MSV       GLKRAGSKADRPSLQIQTLQHAGTTMITVPSGGVCDLINTYARGSDEGNRHTSET..L
WDV       AYKVVPVKPPALCVFRYNWLNSDRTNIVVGNTPRVDLITCFAQGKADNNRHTNQT..V

          =:==:   :::=: :=:::=:===:== =:  =:= := =:: ==:= =:= ::=:  =

MSWDV     ALKRVPSKADRLSLQRYTLLNSGRTNITVPSGGVVDLINTFAQGKDDNNRHTNQT..V


          ddddddddddd    eeeeeeeeeee                  oooooo
STNV      SHKLHVRGTAITVSQTF.RFIW.FR.......DNMNRGTT.PTVLEVLNTANFMSQYN
MSV       TYKIAVD.YHFVADAAACPYSNTGTGVMWLVYDTTPGGQA.PTPQTIFAYPDTLKAW.
WDV       LYKFNIQGTCYMSDASA.PFIGPVRLYHWLVYDAEPK.QAMPDATDIFTMPWNL..L.

          : =:::=== ::::::  :==   =        =:  := : ==: :::: :: ::::

MSWDV     TYKINVQGTHFVADASA.PFINPVRGVMWLVYDATPKGQA.PTATDIFAMPDTLKAW.


           oooo       fffffffffff          gggggggggggggggggg
STNV      PITLQ.....QKRFTILKDVTLNCSLTG.....ESIKDRIINLPG.QLVNYN......
MSV       PATWKVSRELCHRFVVKRRWLFNMETDGR.IGSDIPPSNASWKPCKRNIYFHKFTSGL
WDV       PSTWTVQRAWSHRFVVKRKWTVNLVTDGRKVGSKTVDQRYNWVVGKNIVDANKFFKGL

          = = :      :==:: :: =:= : :=     :::::=   :== ::=::=

MSWDV     PATWKVSRELSHRFVVKRKWTFNMETDGR.IGSKTVDQRASWVPGKNIVDFNKFTSGL


                            hhhhhhhhhh      iiiiiiiiiiiiiiiiiii
STNV      .........GATAVAASNGPGAIFML.QIGDSLVGLWDSSYEAV..........YTDA
MSV       GVRTQWKNVTDGGVGAIQR.GALYMVIAPGNGLTF..TAHGQTRL..........YFKS
WDV       RVTTEWMNTGDGKIGDIKK.GALYLISSTRGGVTG..DSASTAFDVVCAYTHACYFKA

                  =:::=:= :  ==:::=: ::=::=:=  ==:::=          =::=

MSWDV     GVRTQWKNVGDGGVGAIQK.GALYMVIATGNGLTG..DSASQAFL.........YFKA
```

Fig. 1 Alignment of the beta barrel region of STNV (residues 15-195) with WDV (residues 35-222) and MSV (residues 27-214). The lower case letters b to i over the STNV sequence represent the eight strands in STNV that form the beta barrel. The lower case letter o stands for helical regions of STNV. (see ref. 14 for details). Dots in the various sequences represent gaps introduced during the process of sequence alignment. The symbol =(:) shows that the STNV residue at that particular position is identical to (conserved with) at least one of the geminivirus residues. An exchange is considered to be a conserved one if the score for that exchange in the MDM matrix (ref. 17) is greater than or equal to 10.

aligned sequences is well below the random value of 1.45 to 1.50 found in this and other analyses (26). In the case of MDM exchange scores, the average value is above 10.6, the random value being about 9.2 according to this work. The maximum value of the mean parametric correlation coefficient for chance alignment is 0.12 in the present work whereas the average value for 50 random alignments was zero. For all comparisons attempted

Table 1 Statistics for aligned protein pairs

Comparison	Percentage Identical Residues	Conserved	Average value of MBC/C	MDM	Corr. Coeft.	Effective Similarity Score Iden-tity	MBC/C	MDM	Corr Coeft.
MSV:WDV	36	35	0.91	11.97	0.46	18.9	13.5	20.6	11.4
STNV:MSWDV	33	45	0.98	11.72	0.58	14.4	11.0	14.6	12.2
STNV:WDV	24	41	1.13	10.96	0.43	9.4	7.0	10.0	9.0
STNV:MSV	19	45	1.27	10.68	0.34	7.2	5.3	9.3	9.5

The effective similarity score is defined as S(effective) = (S - S(random)) / σ(S) where S is the score for the test alignment, S(random) is the mean score obtained for 50 random alignments and σ(S) is the standard deviation for random aligment scores. For a pair of aligned proteins, keeping the first sequence unaltered, if the second sequence is scrambled, the amino acid composition remains constant for both the proteins, but the sequence alignment beomes random. The scores of identical residues, MBC/C, MDM exchange scores and the mean value of parametric correlation coefficients are found for fifty random alignments. These are averaged and their standard deviations estimated in order to determine the effective similarity score. STNV represents the beta barrel region between residues 15 and 195 (see ref. 24). WDV and MSV represent respectively the regions between residues 35-222 (ref. 9) and 27-214 (ref. 7) in the two viruses omitting the charged amino termini. MSWDV is the consensus sequence of MSV and WDV (see also Fig. 1).

here, the various scores are five to 20 times the standard deviation for the corresponding random values satisfying the criterion of approximately three standard deviations for the threshold of homology (27). STNV aligns better with WDV than with MSV. The STNV:MSWDV alignment is as good as that of the MSV:WDV alignment. If one accepts that MSV and WDV are doubtless homologous, then it follows that these viruses do indeed take the beta barrel topology of STNV. With respect to STNV, there seem to be three prominent insertions in MSV and WDV, two being at the beginning and the end of the region between the two halves of the beta barrel. The third insertion is in a region between the strands β_g and β_h corresponding to STNV. Such an insertion between strands occurs in the structure of SBMV too. These insertions probably have a role to play either in protein-DNA interaction or in the interaction at the interface of the incomplete icosahedra. Even if answers to these and other questions must await a detailed crystal structure analysis of a geminivirus, it is clear that the beta barrel topology is a prerequisite for icosahedral assemblage and transcends genomic preferences.

ACKNOWLEDGEMENTS

The author is grateful to Professor Patrick Argos for inspiration and to Professor Mark A. Hermodson for support. He also thanks Professor Michael G. Rossmann and the Purdue University for computing facilities.

REFERENCES

1. Goodman, R. M. (1981) J. Gen. Virol. 54, 9-21.
2. Howell, S. H. (1985) CRC Crit. Rev. in Plant Sci. 2, 287-316.
3. Harrison, B. D. (1985) Ann. Rev. Phytopathol. 23, 55-82.
4. Stanley, J. and Gay, M. R. (1983) Nature(London) 301, 260-262.
5. Hamilton, W. D. O., Stein, V. E., Coutts, R. H. A. and Buck, K. W. (1984) EMBO J. 3, 2197-2205.
6. Howarth A. J., Caton, J., Bossert, M. and Goodman, R. M. (1985) Proc. Natl. Acad. Sci. 82, 3572-3576.
7. Howell, S. H. (1984) Nucl. Acids Res. 12, 7359-7375.
8. Mullineaux, P. M., Donson, J., Morris-Krsinich, B. A. M., Boulton, M. I. and Davies, J. W. (1984) EMBO J. 3, 3063-3068.
9. MacDowell, S. W., Macdonald, H., Hamilton, W. D. O., Coutts, R. H. A. and Buck, K. W. (1985) EMBO J. 4, 2173-2180.
10. Kikuno, R., Toh, H., Hayashida, H. and Miyata, T. (1984) Nature(London) 308, 562.
11. Mohana Rao, J. K. (1985) Biochem. Biophys. Res. Commun. 130, 892-896.
12. Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. and Bricogne, G. (1978) Nature(London) 276, 368-373.
13. Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G. W., Rayment, I., Rossmann, M. G., Suck, D. and Tsukihara, T. (1980) Nature(London) 286, 33-39.
14. Liljas, L., Unge, T., Jones, T. A., Fridborg, K., Lovgren, S., Skoglund, U. and Strandberg, B. (1982) J. Mol. Biol. 159, 93-108.
15. Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H-J., Johnson, J. E., Kamer, G., Luo, M., Mosser, A. G., Rueckert, R. R., Sherry, B. and Vriend, G. (1985) Nature(London) 317, 145-153.
16. Hogle, J. M., Chow, M. and Filman, D. J. (1985) Science, 229, 1358-1365.
17. Dayhoff, M. O., Scwartz, R. M. and Orcutt, B. C. (1978) in Atlas of Protein Sequence and Structure, Vol. 5, Suppl. 3, Natl. Biomed. Res. Foundn., Washington, D.C., pp 345-352.
18. Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 13, 211-221.
19. Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 13, 222-245.
20. Palau, J., Argos, P. and Puigdomenech, P. (1982) Int. J. Peptide Prot. Res. 19, 394-401.
21. Jones, K. K. (1975) J. Theor. Biol. 50, 167-183.
22. Sweet, R. M. and Eisenberg, D. (1983) J. Mol. Biol. 171, 479-488.
23. Zalkin, H., Argos, P., Narayana, S. V. L., Tiedman, A. A. and Smith, J. M. (1985) J. Biol. Chem. 260, 3350-3354.
24. Henriksson, D., Tanis, R. J., Tashian, R. E. and Nyman, P. O. (1981) J. Mol. Biol. 152, 171-179.
25. Hopper, P., Harrison, S. C. and Sauer, R. T. (1984) J. Mol. Biol. 177, 701-713.
26. Keim, P., Heinrikson, R. L. and Fitch, W. M. (1981) J. Mol. Biol. 151, 179-197.
27. Feng, D. F., Johnson, M. S. and Doolittle, R. F. (1985) J. Mol. Evoln. (1985) 21, 112-125.