Vol. 83, No. 3, 1978 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 14, 1978 Pages 863-868

POLYCONFIGURON-MODEL FOR THE A-PROTEIN OF COLIPHAGE MS2

Loyd Y. Quinn

Department of Bacteriology
Iowa State University
Ames, Iowa 50011

Received June 14,1978

SUMMARY: An algorithm is presented for decoding mRNA to give the tertiary structure of A-protein of MS2 coliphage. The scale model for the A-protein was assembled by stepwise "translation" of each codon into the corresponding amino acid configuron, so that a polyconfiguron composed of the 393 amino acid residues of the A-protein was generated. This polyconfiguron has the topography of native A-protein, as it has been predicted by a variety of functional and physicochemical procedures.

Group I RNA coliphage has been studied extensively since it "represents a biological system uniquely reduced to its simplest form. Unlike all other bacterial viruses, the RNA phage does not contain DNA, but rather RNA as its sole genetic material. As a consequence, the RNA must serve a dual function both as a template for nucleic acid synthesis and as a messenger for virus-specific protein synthesis. Due to the size of the genome and the limited number (three) of gene products, it has been possible to elucidate the biological processes of replication and translation as well as the mechanisms by which these events are controlled---" (1). MS2 is the best-characterized of the Group I RNA coliphage, and is the first biological system for which the entire nucleotide sequence of the genome and the amino acid sequences of all protein products of the genome have been determined (2). The three genes of these phages encode for the A-protein, coat protein monomer, and RNA replicase polypeptide β, respectively. Each of these proteins has been shown to be poly-functional (2-17). The A-protein has been compared to the gene H spike protein of ØX174 phage (18, 19) which acts as a "pilot protein" with multiple functions in adsorption, penetration, and early intracellular stages of viral chromosome expression. However, the A-protein also functions as maturation protein, and as lysozyme for release of mature phage from infected E. coli

0006-291X/78/0833-0863$01.00/0

Hfr cells. During assembly of the MS2 virion the RNA binds the three proteins, coils into an α-helix, and acts as template for assembly of the icosahedral capsid composed of 180 coat protein monomers. Amber mutants of the A-protein gene give rise to MS2 virions that lack normal lysozyme activity, or fail to bind to the F-pilus, or whose RNA is attacked by RNase.

The ribosomal microcycle in E. coli requires about 50 microseconds, so that the A-protein polypeptide chain of 393 amino acid residues may be translated from its mRNA and released from the ribosome in native configuration in about 20 seconds. Since even a polypeptide of 150 amino acid residues would require an estimated 10^{26} years to reach native configuration by thermodynamic processes, Anfinsen and Scheraga (20) have proposed that "nucleation sites" form in the nascent polypeptide by short range interactions between the side chain and the atoms of the backbone of the same amino acid. These nucleation sites then serve as templates for rapid assembly of the polypeptide chain into a native configuration with lowest free energy. The 30-35 most proximal amino acid residues of the nascent polypeptide are held within the ribosomal groove (21, 22) where nucleation sites may be completed before the polypeptide enters the destabilizing aqueous environment of extra-ribosomal space. The tertiary structures of conformational antigenic sites are generated before the polypeptide chain is completed and released from the ribosome (23), which points to a stepwise generation of tertiary structure of proteins during mRNA translation.

Key requirements for this stepwise genetic control of protein tertiary structure during mRNA translation are mechanisms for preserving nearly perfect homology between each mRNA codon and the corresponding anticodon of cognate aminoacylated tRNA, coupled with dynamic template functions that stereospecifically insert nascent amino acid residues. The tRNA synthetases insure high fidelity of translation, as shown by Yamane and Hopfield (30), by kinetic proofreading during aminoacylation that discriminates against non-cognate amino acids and tRNAs. Lake (31) has used the complete map of proteins of the

ribosomal small subunit to propose a new model for ribosome-tRNA interactions: The first step of the elongation cycle of protein synthesis is binding of aminoacyl-tRNA to the R (recognition) site during codon recognition. Next, a conformational change in the anticodon loop from the 5' stacked conformation to the 3' stacked conformation moves the acceptor end of the tRNA from the R site to the A site, but during this process the messenger and the anticodon remain fixed on the ribosomal surface. This brings the acceptor end of the aminoacyl-tRNA and the acceptor end of the peptidyl-tRNA (occupying the P site) into adjacent positions and allows recognition of the correct reading frame by a base-pairing interaction between A and P site tRNAs through *trans* pairing of the invariant bases U-33 of both molecules.

The *polyconfiguron* model for decoding mRNA into the tertiary structure of the corresponding protein proposes that the movement of the CCA end of aminoacylated tRNA, in the switch from the R site to the A site, causes the sidechain of the bound amino acid to assume a configuration that has the greatest cooperativity with all of the other participants in the translation process. This paper describes a test of the polyconfiguron algorithm for mRNA decoding, using the A-protein of MS2 coliphage. The A-protein was selected for this purpose because it (a) has known primary structure, (b) has known mRNA codon sequence, (c) has been partially characterized both functionally, and in regard to physico-chemical properties, and (d) does not undergo posttranslational modification.

MATERIALS AND METHODS

The polyconfiguron algorithm for decoding mRNA is derived as follows:

1. Convert trinucleotide codons to binary arithmetic numbers that preserve both the format and the complementary-base-pairing patterns implicit in codons, by setting A=01, U=10, G=00, and C=11.

2. Convert the binary numbers to magnetic dipole equivalents by representing each "0" bit as S and each "1" bit as N of bar magnets mounted at midpoints on a common axis about which they are free to rotate. Measure the deflections of each magnet as counterclockwise rotational angles from the vertical reference point.

3. Mount a Kendrew-type (24) model of glycine on a Zeiss Universal Stage, with the amino acid in the standard nonrotated position (25), in which the

TABLE I. Algorithm for Conversion of Codons to Configurons.

Configuron	RNA Codon	Binary Number	Magnetic Patterns	Clockwise Rotational Angles					
				Bit 6	Bit 5	Bit 4	Bit 3	Bit 2	Bit 1
Gly-1	GGG	000000	SSSSSS	80	266	96	289	114	298
Gly-2	GGA	000001	SSSSSN	250	263	99	285	108	292
Gly-3	GGU	000010	SSSSNS	53	121	137	322	150	335
Gly-4	GGC	000011	SSSSNN	248	74	89	281	106	290

model lies in the horizontal plane with the amino group to the left on the x-axis, and with the sidechain (H for glycine) to the right, representing the natural L-isomer. For each codon, apply the first three angles to rotate the amino group counterclockwise about the α-carbon, in three mutually perpendicular planes; apply the last three angles of each set to rotate the carboxyl group counterclockwise about the α-carbon in three mutually perpendicular planes. the derivation of this data for the glycine codons is indicated in Table 1.

RESULTS AND DISUCSSION

This operation was carried out for each of the 64 mRNA codons, thereby producing the corresponding set of 64 amino acid configurons, which represent the 3-D configurations of cognate amino acids at the moment of translation.

"Translation" of the A-protein mRNA into the A-protein Polyconfiguron:

Kendrew-type wire models of required amino acids were built to 1:2 x 10^8 scale, the amino acids were oriented in the configuron mode specified by the corresponding codon, and were joined by peptide linkage. The backbone structure of the resulting polyconfiguron model of the A-protein is shown in Figure 1, where every 10th α-carbon was included in the folding pattern.

The polyconfiguron may be described as a divergent pair of pleated filaments leading into a globular region which is composed of two semicircular bends of approximately 48 A diameter, very near the diameter of an apical penton of the MS2 capsid. The inner margins of the semicircular bends are lined with 10 basic amino acid residues and form a hole approximately 20 A in diameter, which is the same as the diameter of the α-helix of RNA; this region has been proposed as the binding site for MS2 RNA (2). The bottom surface of the semicircular bends is composed of 9 hydrophobic amino acid residues plus the three cysteine residues; these "sticky" residues may well form the specific binding site for the F-pilus, which has a similar curvature as this region.

The A-protein polyconfiguron is stabilized by 7 salt bridges, 2 very short α-helices, 7 pi-bonding areas, 100 β-bonds and over 300 hydrophobic interactions. In addition, three potential Ca-chelating sites are present, which may explain the need for calcium ions in the infection process. Immunogenic amino acid residues, namely Tyr, Lys, Glu, Phe, and Ala (28) showed a strong tendency to cluster at five sites on the A-protein polyconfiguron. One especially interesting immunogenetic site contains the sequence 161Val.162Lys. 163Val, which is also present in the immunodominant site of the MS2 coat monomer (29); this probably explains the cross-reactivity we have observed between Anti-A-protein and anti-coat protein antibodies, both of which inhibit infection of E. coli Hfr by MS2 phage.

The scale model polyconfiguron of the A-protein of MS2 coliphage was found to fit well with immunochemical and biophysical models previously described

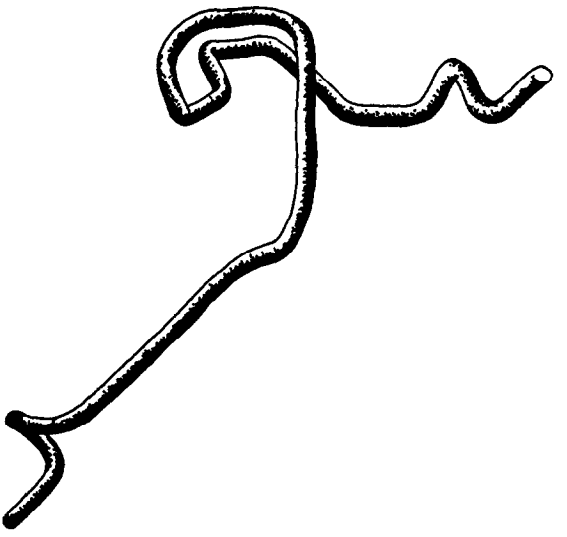

Figure I. Backbone structure of the A-protein Polyconfiguron. View of the molecule from the amino terminus, lower left, to carboxyl terminus, upper right. Bends were checked at every tenth α-carbon along the chain. The configuration fits well with A-protein functions such as binding the RNA α-helix, binding to the F-pilus, and forming an operculum for the capsid.

for this protein. In addition, the Chou and Fasman algorithims (26, 27), which are 77% accurate in predicting the backbone structure of 29 well-defined proteins, predict the almost exclusively β-sheet structure found for this polyconfiguron.

ACKNOWLEDGEMENTS

Thanks is expressed to Mark Hartman for the drawing of the backbone structure of the A-protein, and to Terrence Quinn for building most of the Kendrew scale mode amino acids used in this polyconfiguron model.

REFERENCES

1. Eoyang, L., and August, J. T. (1974) Comprehensive Virology 2, pp. 1-59, Plenum Press.

2. Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G., and Ysebaert, M. (1976) Nature 260,500-507.

3. Bonner, Philip H. (1974) J. Virology 14,1152-1168.

4. Crowther, R. A., Amos, L. A., and Finch, J. T. (1975) J. Mol. Biol. 98, 631-635.

5. Curtiss, L. K., and Krueger, R. G. (1974) J. Virol. 14,503-508.

6. Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Merregaert, J., Min Jou, W., Raeymaekers, A., Volckaert, G., and Ysebaert, M. (1975) Nature 256,273-378.

7. Jenkins, S. T., and Beard, J. P., and Howe, T. G. B. (1974) J. Virol. 14,50-55.

8. Luria, S. E., Darnell, J. E., Baltimore, D., and Campbell, A. (1978) General Virology, 3rd ed., pp. 205-209.

9. O'Callaghan, R., Bradley, R., and Paranchych, W. (1973) Virology 54, 476-494.

10. Reynolds, S., and Paranchych, W. (1976) Can. J. Microbiol. 22,1647-1653.

11. Rohrmann, G. E., and Krueger, R. G. (1970) J. Virol. 6,269-279.

12. Rymar, S. Yu, and Kordyum, V. A. (1976) Mikrobiol. Zh. 38,552-556.

13. Schoulaker, R., and Engelberg-Kulka, H. (1978) J. Virol. 25,433-435.

14. Shiba, T., and Miyake, T. (1975) Nature 254,157-158.

15. Thomas, G. J., Prescott, B., McDonald-Ordzie, P. E., and Hartman, K. A. (1976) J. Mol. Biol. 102,103-124.

16. Tikhonenko, A. S., Bespalova, I. A., and Kriviskii, A. S. (1965) Dokl. Akad. Nauk SSSR 160,704-706.

17. Turchinsky, M. F., Kusova, K. S., and Budowsky, B. I. (1974) FEBS Letters 38,304-307.

18. Jazwinski, S. M., Lindberg, A. A., and Kornberg, A. (1975) Virology 66,283-293.

19. Jazwinski, S. M., Marco, R., and Kornberg, A. (1975) Virology 66,294-305.

20. Anfinsen, C. B., and Scheraga, H. A. (1975) Advances in Protein Chemistry, Vol. 29, pp. 205-300, Academic Press, N.Y.

21. Kurland, C. G., Rigler, R., Ehrenberg, M., and Blomberg, C. (1975) PNAS 72,4248-4251.

22. Lake, J. A. (1976) J. Mol. Biol. 105,131-159.

23. Schechter, I. (1974) Biochem. 13,1876-1885.

24. Kendrew, J. C. (1961) Scientific American 205,96-121.

25. Dickerson, R. E., and Geis, I. (1969) The Structure and Action of Proteins, pp. 1-23, Harper and Row, Publishers, N.Y.

26. Chou, P. Y., and Fasman, G. D. (1974) Biochem. 13,222-245.

27. Fasman, G. D., Chou, P. Y., and Adler, A. J. (1976) Biophys. J. 16,1201-1238.

28. Gill, T. J. (1973) Specific Receptors of Antibodies, Antigens and Cells. pp. 136-169, Karger, Basel.

29. Langbeheim, H., Arnon, R., and Sela, M. (1976) PNAS 73,4636-4640.

30. Yamane, T., and Hopfield, J. J. (1977) PNAS 74,2246-2250.

31. Lake, J. A. (1977) PNAS 74,1903-1907.