Vol. 97, No. 3, 1980 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
December 16, 1980 Pages 868-874

CONFIGURON DEPENDENCE ON TRANSLATION OF SPECIFIC CODON PAIRS. I. HELICAL REGIONS OF HUMAN ALPHA AND BETA GLOBINS.

Loyd Y. Quinn[1], Ramachandran Krishnaswamy[2], and Terrence R. Quinn[3]

Departments of Microbiology[1] and Computer Science[2], Iowa State University, and Department of Mathematics[3], University of Colorado, Boulder, Colorado

Received September 20,1980

SUMMARY: Homologous proteins, which possess similar shapes, functions, and amino acid sequences, are encoded by homologous messenger ribonucleic acids whose codon sequences tend to be similar. It is proposed that helical configurons are generated when certain pairs of contiguous codons are translated, and that non-helical configurons appear when other specific pairs of codons are read off. The resulting sequence of configurons comprises the polyconfiguron, which forms the native structure of the protein.

Globular proteins of similar function, regardless of species of origin, tend to be homologous with respect to amino acid sequences (1-5). In those homologous globular proteins that have been defined in terms of molecular configurations, amino acid sequences, and mRNA sequences, it also has been shown that this homology extends to specific mRNA sequences (6-11). This relationship between specific mRNA codon sequences and the native shape of homologous proteins has been interpreted by the configuron model for mRNA translation (12), which proposed that specific mRNA codon sequences are read directly into stereospecific configurations of cognate amino acids, the configurons, through the mediation of aminoacyl-tRNAs and ribosomal translation sites.

Recently reported observations of tRNA-tRNA interactions, both in association with aminoacyl-tRNA synthetases and with the A-site/P-site regions of the ribosome (14-23) imply that mRNA codons are translated as contiguous pairs, rather than singly; in light of these findings, the polyconfiguron model has been expanded to propose that there is direct genetic control of the configuration of nascent peptides ("dipeptides") as each contiguous pair of mRNA codons is translated. It is assumed that each amino acid is in most

0006-291X/80/230868-07$01.00/0

stable configuration (24-26) when bound to cognate tRNA; this is suggested by the report of Gelin and Karplus (27) who found that the minimal energy positions of amino acid residues in stable regions of bovine pancreatic trypsin inhibitor were close to those of the corresponding free amino acids. It would seem that secondary structure is controlled stepwise by the manner in which the flexible aminoacyl-tRNA:aminoacyl-tRNA pairs interact as they present the relatively rigid amino acid pairs to the peptidization site on the ribosome, during mRNA translation.

This postulated association between contiguous pairs of codons and corresponding configurations of cognate peptides was explored as follows: (a) X-ray crystallographic data for each amino acid residue in a selected protein was converted into the corresponding configuration by rotating the amino acid molecule counterclockwise around its alpha-carbon, in three mutually perpendicular planes. The coordinates of the atoms for the preceding amino acid residue were used as the starting points for these configuron rotations (CR). The Configuron computer program for this operation was written in FORTRAN IV, and is available from the authors; (b) The configuron produced by a given set of paired codons in cognate mRNA was then compared with the configuration of this residue determined by other methods (29-32).

MATERIALS AND METHODS

Selection of Suitably Defined Proteins for Configuron Analysis

Since, in only structurally stable regions of proteins is it possible to test for paired-codon control of configurons, a method for locating stable regions of crystalline proteins had to be devised. This consisted of determining which amino acid residues showed no more than 10° variance in dihedral angles attributable to thermal fluctuations (33) between sets of crystallographic data for the same protein filed with the Protein Data Bank (34) by different depositors. Human globin-A and globin-B were the proteins used in this study because they are the only x-ray defined proteins (35, 36) for which both the amino acid sequences (37-40) and corresponding mRNA codon sequences (7-8) have been reported.

In the A-chain the following regions were found to be stable, by residue numbers: 3-4,7-14,18,21-23,25-26,30-40,42-45,48-49,51,55-58,60-65,67-68, 70,74-76,78-82,84,86-87,89-90,93-95,100-101,103-104,106-107,109-113,118-121, 123-134, and 136-138.

TABLE I. Relationships of Paired Contiguous Codons To Configurons in Helical Regions of Human Globins A and B

Codon-Pair	Chain	Residue-Pair	File 71			File 73		
			CR1	CR2	CR3	CR1	CR2	CR3
GCC:GCC	A	12ALA:13ALA	61	10	289	49	19	280
GCC:GCC	A	110ALA:111ALA	68	7	298	48	19	302
GCC:GCC	B	128ALA:129ALA	57	12	287	67	8	287
CUG:CUG	A	105LEU:106LEU	56	23	280	70	6	294
CUG:CUG	B	31LEU:32LEU	69	13	290	55	17	297
UUC:CUG	A	33PHE:34LEU	63	7	299	57	17	291
UUC:CUG	A	128PHE:129LEU	63	12	291	57	16	298
CUG:CAC	A	86LEU:87HIS	58	10	300	64	7	303
CUG:CAC	B	91LEU:92HIS	65	8	292	58	9	304
CUG:GUG	A	106LEU:107VAL	71	12	293	68	5	285
CUG:GUG	B	32LEU:33VAL	65	6	298	58	18	286

In the B-chain the stable sequences were 3,6,9-13,16-17,19-20,27-42, 47-48,52-58,60,63,67,72-74,78-81,83,85-93,97-99,103-104,106,108-109,111-113, 115,117-120,123-126,129-130,132-138,141,143, and 146.

It may be noted that these stable regions are located in or near the helical sites in these globin chains, where the Phi angles lie in the range of -48° to -67°, Psi angles lie in the range -44° to -57°, and the Omega angle is -180°. This corresponds to Configuron Rotations (CR) of: CR1 = -48° to -66°; CR2 = -11° to -26°; and CR3 = -280° to -300°.

RESULTS

The results of the comparison of paired-codons with configurons are presented in Table I. File 71 and File 73 refer to the Protein Data Bank 1980 magnetic tape file numbers for the crystallographic data of Fermi, and of Baldwin and Chothia, respectively.

Relationships of Paired-Codons to Configurons in Non-Helical Regions

As indicated earlier, the non-helical regions of human globins A and B are too labile to be well-defined by x-ray crystallographic analysis, but certain trends in relationships between paired-codons and corresponding configurons are evident. For example, the strong helix-forming residues such as ALA, GLU, LEU, and MET (31-32) can occur either in helices or in β-turns, depending on the nature of the corresponding encoding codon-pairs. Likewise, the strong helix-breakers such as GLY and PRO (31-32) can occur either in helices or in non-helical configurations, as illustrated by these examples: (a) in the C-helix of the A-chain 36PHE.UUC:37PRO.CCC gives to the configuron for PRO of CR = -81°/-61°/-156° which correspondes to a right-handed β-turn; (b) in the H-helix of the B-chain 124PRO.CCA:125PRO.CCA gives rise to the configuron for PRO of CR = -62°/-3°/-299° which corresponds to the α-helix. In similar fashion, ALA.GCU: GLY.GGC, and LYS.AAG:GLY.GGC generate GLY configurons of helix-forming type, but PHE.UUU:GLY.GGC gives rise to a GLY configuron of right-handed β-turn type.

Interpretation of Configuron Behavior by Scale Model Building

By inspection of Kendrew-type wire models of proteins (41) built by assembly of the specified configurons into a polyconfiguron, it is apparent that: (a) during chain-elongation, the C-terminal configuron is inverted with respect to the configuration of the penultimate configuron, a mechanism that insures preservation of the normal trans configuration of the α-carbons with respect to the peptide bond; and (b) within the limits of the data, it appears that the same amino acid assumes the same stereoisomeric form each time it is incorporated into a given polypeptide chain, regardless of the choice of codon synonym which encodes for it in the corresponding mRNA.

DISCUSSION

Native configurations of labile proteins, such as globins, are subject to denaturation by procedures commonly employed in purification, crystallization, and preparation for x-ray crystallographic analysis. These artifacts tend to occur in the least stable regions of protein chains, such as at termini, or in β-turns, and may be detected by comparison of coordinate maps of a protein produced by separate x-ray analyses. The stable helical core regions of globular proteins seem to be nearly the same in configuration in various crystal forms, with appreciable configurational changes due to different packing effects appearing only in surfaces regions of the protein (42).

In proteins which are produced by mRNA translation, the native configurations develop in stepwise fashion almost simultaneously with chain elongation (43), and the nascent chain is stabilized by ligand-binding and by hydration (44, 45). Numerous studies have indicated that the stable regions of proteins are similar in configuration in crystalline and in aqueous solution (6, 25, 28), a finding that relates form and function in these biologically active molecules.

The configuron model for generation of protein configurations seems to be supported by the limited data available from crystallographic analyses of human globins A and H. It will be useful to apply the configuron model to a protein such as hen egg white lysozyme, which is well-characterized by x-ray crystallography and other appropriate techniques, but for which the complete mRNA sequence has not been reported in the literature. Lysozyme

is of especial interest because it is predominately of β-sheet, rather than helical, configuration, and is a relatively stable molecule (42, 46, 47).

REFERENCES

1. Dayhoff, M. O. (1972) Atlas of Protein Sequence and Structure (Data Section) National Biomedical Research Foundation, Silver Spring, MD.
2. Dayhoff, M. O. (1973) Ibid. Supplement 1.
3. Dayhoff, M. O. (1976) Ibid. Supplement 2.
4. Dayhoff, M. O. (1978) Ibid. Supplement 3.
5. Feldman, R. J. (1976) Atlas of Macromolecular Structure on Microfiche, Tracor Jitco Inc., Rockville, MD.
6. Mathews, B. W. (1976) Ann. Rev. Phys. Chem. 27,493-523.
7. Wilson, J. T., Wilson, L. B., Reddy, V. B., Cavellesco, C., Ghosh, P. K., deRiel, J. K., Forget, B. G. and Weissman, S. M. (1980) J. Biol. Chem. 225,2807-2815.
8. Marotta, C. A. Wilson, J. T., Forget, . G., and Weissman, S. M. (1977) J. Biol. Chem. 252,5040-5053.
9. Heindell, H. C., Liu, A., Paddock, G. V., Studnicka, G. M., and Salser, W. A. (1978) Cell 15,43-54.
10. Richards, R. I., Shine, J., Ullrich, A., Wells, J. R. E., and Goodman, H. M. (1979) Nucl. Acids Res. 7,1137-1146.
11. van Ooyen, A., van den Berg, J., Mantei, N., and Weissman, C. (1979) Science 206, 337-344.
12. Pavlakis, G. N., Lockard, R. E., Vamvakopoulos, N., Rieser, L., RajBhandary, U. L., and Vournakis, J. N. (1980) Cell 19,91-102.
13. Quinn, L. Y. (1978) Biochem. Biophys. Res. Commun. 83,863-868.
14. Lake, J. A. (1977) PNAS 74,1903-1907.
15. Lestienne, P. (1978) J. Theor. Biol. 73,159-180.
16. Lührmann, R., Eckhardt, H., and Stöffler (1979) Nature 280,423-425.
17. Möller, A., Wild, U., Riesner, D., and Gassen, H. G. (1979) PNAS 76,3266-3270.
18. Abdurashidova, G. G., Turchinsky, M. F., and Aslanov, K. A. and Budowsky, E. I. (1979) Nucl. Acids Res. 6,3891-3909.
19. Davanloo, P., Sprinzl, M., and Cramer, F. (1979) Biochem. 18,3189-3199.
20. Peters, M., and Yarus, M. (1979) J. Mol. Biol. 134,471-491.
21. Ofengand, J., Lin, F. L., and Liou, R. (1979) Arch. Biol. M. 12,470-471.

22. Yoo, S. H., and Shive, W. (1979) Biochem. Biophys. Res. Commun. 88,552-558.

23. Kirillov. S. V., Makhno, V. I., and Semenkov, Y. P. (1980) Nucl. Acids Res. 8,183-196.

24. Hodes, Z. I., Némethy, G., and Scheraga, H. A. (1979) Biopolymers 18, 1565-1610.

25. Deslauriers, R., and Smith I. C. P. (1980) Biological Magnetic Resonance, Vol. 2, 243-345, Plenum Press, N.Y.

26. Hamilton, W. C., Frey, M. N., Golič, L., Koetzle, T. F., Lehmann, M. S., and Verbist, J. J. (1972) Steroscopic Atlas of Amino Acid Structures Mat. Res. Bull. 7,1225-1238, Pergamon Press, Inc.

27. Gelin, B. R. and Karplus, M. (1975) PNAS 72,2002-2006.

28. Siegel, J. B., Steinmetz, W. E. and Long, G. L. (1980) Anal. Biochem. 104,160-167.

29. Cerf, R. (1978) PNAS 75,2755-2758.

30. Levitt, M., and Greer, H. (1977) J. Mol. Biol. 144,181-293.

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

32. Chou, P. Y. and Fasman, G. D. (1977) J. Mol. Biol. 115,135-175.

33. Levy, R. M., and Karplus, M. (1979) Biopolymers 18,2465-2495.

34. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) J. Mol. Biol. 112,535-542.

35. Fermi, G. (1975) J. Mol Biol. 97,237-256.

36. Baldwin, J. M. (1980) J. Mol. Biol. 136,103-128.

37. Konigsberg, W. and Hill, R. J. (1962) J. Biol. Chem. 237,3151-3162.

38. Braunitzer, G., Rudloff, V., and Hilschman, N. (1963) Hoppe-Seyler's Z. Physiol. Chem. 331,1-32.

39. Braunitzer, G., Gehring-Mueller, R., Hillschman, N., Hilse, K., Hobom, G., Rudloff, V., and Wittman-Liebold, B. (1961) Hoppe-Seyler's Z. Physiol. Chem. 325,283-286.

40. Konigsberg, W., Goldstein, J., and Hill, R. J. (1963) J. Biol. Chem. 238,2028-2033.

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

42. Moult, J., Yonath, A., Traub, W., Smilansky, A., Podjarny, A., Rabinovich, and Saya, A. (1976) J. Mol. Biol. 100,179-195.

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

44. Hodes, Z. I., Némethy, G., and Scheraga, A. (1979) Biopolymers 18, 1611-1634.

45. Clementi, E., Cavallone, F., and Scordamaglia, R. (1977) JACS 99,5531-5545.

46. Artymiuk, P. J., Blake, C. C. F., and Oatley, S. J. (1979) J. Chem. Phys. 76,813-815.

47. Diamond, R. (1974) J. Mol. Biol. 82,371-391.