Vol. 163, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 30, 1989 Pages 466-470

SELF-COMPLEMENTARY REGIONS IN HUMAN ALBUMIN mRNA ENCODE IMPORTANT STRUCTURAL REGIONS WITHIN THE HUMAN ALBUMIN PROTEIN

Kenneth G. Draper

Antiviral Chemotherapy
Schering-Plough Corporation
60 Orange Street
Bloomfield, NJ 07003

Received July 17, 1989

Summary: An analysis of the human albumin mRNA structure revealed a nonrandom distribution of self-complementary regions within the mRNA. The majority of these self-complementary mRNA stretches encode important structural regions of the human albumin protein. The amino acids contained within these regions of the protein exhibit a high degree of hydrophobic complementarity which could influence local protein conformation and contribute to the biological importance of the protein structures.
© 1989 Academic Press, Inc.

There can be little doubt that local interactions between amino acids are the primary determinant of secondary structure in a protein (1). The relative contribution of a given amino acid to the local hydrophobic character of a peptide sequence has been estimated by Kyte and Doolittle in a manner which combined both empirical and theoretical methodologies (2). Using these values, it has been observed that complementary DNA strands encode amino acids which exhibit complementary hydrophobic characteristics (3). Subsequent work revealed that nucleic acid sequences which are complementary to hormone mRNA sequences encode peptide segments (called complementary peptides) which bind with high affinity to the appropriate target hormones (4-6). The demonstration that amino acid interactions of complementary peptides can be predicted from the analysis of nucleic acid sequence suggests that the secondary and tertiary structure of a protein may be predictable ab initio from an analysis of complementary regions within an mRNA molecule.

To test this hypothesis, human albumin (HA) mRNA structure was analyzed and compared to the structure of human albumin protein. Self complementary regions were non-randomly distributed within the mRNA. Further analysis revealed that most of the self-complementary sequences encoded regions of amino acid self-complementarity, which were located in structurally important regions of the HA protein. The contribution of

0006-291X/89 $1.50
Copyright © 1989 by Academic Press, Inc.
All rights of reproduction in any form reserved.

amino acid hydrophobicity to the proposed secondary structure of one of these self-complementary regions is illustrated.

METHODS

The analysis of the human albumin mRNA sequence (7) was performed using the IBI Pustell Sequence Analysis Software.

The structure of the human albumin protein was proposed previously by Ledden and Feldhoff (8). This model was used to determine the location of specific amino acid chains and protein turns within the human albumin molecule.

RESULTS AND DISCUSSION

Analysis of HA mRNA sequence

The protein coding region of the HA mRNA sequence was searched for examples of extended nucleotide complementarity. To filter the noise associated with this search, regions of homology were limited to those which extended over windows of greater than 30 nucleotides. Linear regions of self-complementary nucleotide sequence were found to be clustered within nucleotides 1-637 and 1212-1830 of the mRNA molecule. These regions of condensed mRNA structure were separated by a stretch of nucleotides which contained only four regions of self complementarity (Table 1).

The tripartite pattern of HA mRNA structure strongly resembles the structure of HA protein which was predicted from hydrodynamic studies (9). This analysis of HA structure suggested that each molecule consists of three spheres in a row with diameters of approximately 38, 53, and 38 A. Interestingly, the three spheres of the HA protein contain approximately equal numbers of amino acids (8). Thus, like its mRNA, HA protein is also composed of two regions of condensed structure which are separated by a domain of more extended structure. These observations suggest that self-complementary structure of the HA mRNA may correlate with a biologically important structural feature of the HA protein.

Structure localization of amino acids encoded by self-complementary regions of HA mRNA

To determine the significance of the mRNA structure to HA protein structure, amino acid stretches encoded by self-complementary mRNA sequences were mapped to locations within the HA protein. Using the structural model of Ledden and Feldhoff (8), fifteen of the twenty-three amino acid stretches were localized to turn regions of the HA protein. Of the remaining amino acid stretches, three contained cysteine (C) residues

Table 1. Location of self-complementarity within human albumin mRNA

Protein Domain	Complementary nucleotides	Amino acids encoded by mRNA sequence	Structural feature
—	54-84	RGVFRRDAHK	proalbumin cleavage site
I.	157-190	QYLQQCPFEDH	turn
	182-229	EDHVKLVNEVTEFAKT	abuts C-C bond
	227-259	CVADESAENC	turn
	248-284	ENCDKSLHTLFG	C-C bond
	275-334	LFGDKLCTVATLRETYGEM	turn
	439-472	MCTAFHDNEET	turn
	468-501	TFLKKYLYEIA	??
	550-608	RYKAAFTECCQAADKAACL	turn
	603-637	LLPKLDELRDE	turn
II.	730-772	SQRFPKAEFAEVSK	turn
	1040-1089	DVFLGMFLYEYARR	turn
	1065-1101	YEYARRHPDYS	turn
	1199-1230	FKPLVEEPQN	turn
III.	1212-1262	VEEPQNLIQNCELFE	turn
	1244-1280	CELFEQLGEYN	C-C bond
	1333-1372	PTLVEVSRNLGKV	??
	1386-1431	KHPEAKRMPCAEDYL	turn
	1472-1502	TPVSDRVTKC	turn
	1535-1588	FSALEVDETYVPKEFNAE	turn
	1566-1605	PKEFNAETFTFHAD	??
	1682-1735	PKATKEQLKAVMDDFAAF	turn
	1715-1750	MDDFAAFVEKCC	C-C bond

Nucleotide numbering begins at the AUG codon where translation of unprocessed albumin protein is initiated. Standard one-letter amino acid code is used. The data is divided into sections which represent the three globular regions of the human albumin protein. C-C bond = cysteine-cysteine bond.

which are known to participate in interstrand C-C bonds, two abut sites of C-C bonds, and another was localized to the site of proalbumin processing.

Analysis of amino acid complementarity in a protein loop region

An analysis of the contributions of complementary amino acid interactions to human albumin protein structure is complicated by the presence of C-C bonds at or near many turns in the molecule. One proposed loop in the protein which should be free of C-C bond influence is located

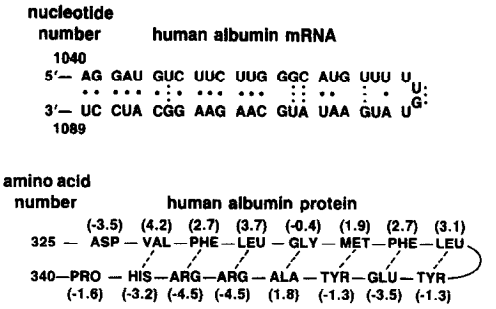

Figure 1. Proposed self-complementary nature of a region of human albumin mRNA and its encoded peptide fragment. The sequence of HA mRNA (7) between nucleotides 1040 and 1089 is shown. Watson-Crick base pairing is represented by a heavy dot. Possible non-Watson-Crick base pairs are shown as three light dots. The putative loop structure of the encoded peptide is shown. This structure is maximized to include the most stable hydrophobic interactions within the peptide region.

between amino acid residues 325-345. This region was chosen for a detailed analysis of amino acid complementarity, because it is also encoded by a self-complementary region of mRNA. An analysis of amino acid content predicts that this region of the protein contains a turn which is stabilized by strong hydrophobic interactions between neighboring amino acids (Figure 1). The proposed hydrophobic interactions within this region were not verified experimentally, but previous reports of the unilateral distribution of hydrophobic residues in regions of turns are consistent with the idea that hydrophobic interactions stabilize turns in proteins.

Various distributions of hydrophilic and hydrophobic residues were observed among the amino acid regions encoded by self-complementary in mRNA sequences. Each region exhibited a center of symmetry which was flanked by peptide segments of short, alternating runs of hydrophobic or hydrophilic residues. Amino acid complementarity between the opposing peptide strands was absent from only two of the amino acid regions listed in Table 1. In both of these instances, the region of nucleotide complementarity overlapped with a larger region of nucleotide complementarity which did encode peptide regions of hydrophobically complementary amino acids.

In summary, a strong correlation exists between the sites of nucleotide complementarity within human albumin mRNA and important structural features of the human albumin protein. The biological implications of this correlation are unclear, but it is tempting to speculate a role of mRNA structure in controlling the rate of translation, which in turn would control the rate of protein folding by allowing interaction of the hydrophobically complementary amino acids encoded by the RNA molecule. The ultimate prediction of protein structure ab initio from amino acid or mRNA sequence will require the ability to predict nucleation sites for protein folding. The algorithms for these predictions should be derived from empirical observations of many turn structures.

Presently, the mRNAs of a number of globular proteins are being analyzed to determine the frequency of occurrence of self-complementary structures and the correlation of these structures to corresponding turns in proteins.

ACKNOWLEDGMENT

I heartily thank Florence Hatchett for the expert typing of the manuscript.

REFERENCES

1. Schultz, G.E., Barry, C.D., Freidman, J., Chou, P.Y., Fasman, G.D., Finkelstein, A.V., Lim, V.I., Ptitsyn, O.B., Kabat, E.A., Wu, T.T., Levitt, M., Robson, B., and Naganao, K. (1974). Nature 250, 140-142.
2. Kyte, J. and Doolittle, R.F. (1982). J. Mol. Biol. 157, 105-132.
3. Blalock, J.E. and Smith, E.M. (1984). Bioch. Biophys. Res. Commun. 121, 203-207.
4. Bost, K.L., Smith, E.M. and Blalock, J.E. (1985). Proc. Natl. Acad. Sci. USA 82, 1372-1375.
5. Bost, K.L. and Blalock, J.E. (1986). Molec. Cell. Endocrinol. 44, 1-9.
6. Mulcahey, J.J., Neill, J.D., Dion, L.D., Bost, K.L., and Blalock, J.E. (1986). Proc. Natl. Acad. Sci. USA 83, 9714-9718.
7. Dugaiczyk, A., Law, S.W. and Dennison, O.E. (1982). Proc. Natl. Acad. Sci. USA 79, 71-75.
8. Ledden, D.J. and Feldhoff, R.C. (1983). J. Protein Chem. 2, 303-319.
9. Bloomfield, V. (1966). Biochemistry 5, 684-689.