56 ml (peak III). FB extracts showed 3 similar peaks with elution volumes of 16, 35 and 52 ml respectively. Comparison of the peaks with the elution volumes of known proteins by the method of Andrews[6] showed the SB extract to contain proteins with mol.wt of 158,500 (peak I); 12,590 (peak II) and also material of mol.wt 1585 for peak III. The FB extract contained proteins of mol.wt 177,800, 22,390 and 3162 for peaks I, II and III respectively. Peaks I and II for each extract were not only inhibitory (figure) to local lesion production but could also be separated from peak III by precipitation on adding ethanol to 80% concentration or by using ammonium sulphate. The material in peak III for each extract enhanced TNV local lesion production and appears to be, at least in part, proteinaceous and clearly different from the oxalate augmenter desribed by Benda[7] and Matsashita[8]. The inhibitor fractions of peak I are of much larger molecular weight than previously described plant virus inhibitors[9,10] and resemble more closely lectins which have been extracted from both SB and FB seeds. SB lectin (mol.wt 110,000) and phytohaemagglutinin (PHA), the lectin from FB (mol.wt 128,000), exhibit a wide variety of biological activities including effects on cell surfaces by binding specifically to carbohydrate-containing receptors on membranes[12]. SB lectin and PHA have affinities for N-acetyl-D-galactosamine although SB additionally binds to D-galactose. Initial stages of plant virus infection involve the entry of virus through the plasmalemma with subsequent multiplication of virus particles. Lectins, by attachment to specific carbohydrate-containing regions of the membrane, may block or modify infectible sites so preventing virus entry. Inhibitors of animal virus multiplication by plant lectins comes about, at least partly, by effects on cell surfaces[11]. The seed extracts examined and described in this paper may contain lectin or lectin-like proteins and their virus inhibitor properties may be related to their effects on cell membranes.

Although the rôle of lectins in plants is not understood it has been suggested that they protect plants against diseases induced by bacteria and fungi[13]. Further studies of SB and FB seed extracts and their authentic lectins should help to verify the possible rôle of these proteins in plant defence against disease.

1 B. Hajj, Thesis, University of London, London 1976.
2 W.A. Stevens, Experientia *26*, 1268 (1970).
3 B. Kassanis, Description of Plant Viruses No. 14. Commonwealth Mycological Institute, London (1970.
4 A. Kleczkowski, Ann. appl. Biol. *36*, 139 (1949).
5 F.C. Bawden, Adv. Virus Res. *2*, 31 (1954).
6 P. Andrews, Biochem. J. *91*, 222 (1964).
7 G.T.A. Benda, Virology *2*, 438 (1956).
8 K. Matsushita and M. Sanada, Physiol. Pl. Path. *13*, 365 (1978).
9 G.E. Apablaza and C.C. Bernier, Can. J. Bot. *50*, 1473 (1972).
10 T. Taniguchi and T. Goto, Ann. phytopath. Soc. Japan *42*, 42 (1976).
11 M.S. Finkelstein and M. McWilliams, Virology *69*, 570 (1976).
12 J.C. Brown and R.C. Hunt, Int. Rev. Cytol. *52*, 277 (1978).
13 I.E. Liener, A. Rev. Pl. Physiol. *27*, 291 (1976).

Base pairing in messenger RNA's for small peptides

D.A.B. Young, R.C. Mann and C.R. Beddell

Wellcome Research Laboratories, Beckenham (Kent BR3 3BS, England), 27 February 1979

Summary. Longest runs of Watson-Crick pairing in hypothetical m-RNA's for a number of natural peptides were no greater than those in the hypothetical m-RNA's for a large number of randomized amino acid sequences from these peptides. This shows that even if base-pairing in m-RNA were a biological requirement, it would little constrain the amino acid sequence.

An interesting question in the fundamental biology of proteins is what, if any, are the a priori restrictions on the possible amino acid sequences. One such restriction could be the relative instability of some messenger RNA's. The stability of these very long chain molecules would be enormously increased by folding stabilized by Watson-Crick pairing of the bases. This would require many regions of contiguous base pairs. The published base sequences for m-RNA's show the possibilities of stems containing short regions of contiguous pairing, but the information is too limited to be an adequate test. The amino acid sequences of many proteins and peptides are known and from these the possible base sequences of their respective m-RNA's may be conjectured, the true sequence being only one amongst many because of the redundancy in the code.

We had preliminary evidence favouring such pairing from the hypothetical sequences for angiotensinogen, when Polya[1] indicated for somatostatin a sequence of bases that can be bent to allow extensive base-pairing. However, there were 2 deficiencies in these examples: a) only one bending point along the RNA chain was examined; and b) probabilities would be very difficult, if not impossible, to calculate.

We have remedied these deficiencies a) by using a computer programme that generates all possible base sequences for a peptide and examines the runs of contiguous pairs for all possible positions of bending; and b) by basing probabilities on the comparison of the data obtained for the peptide with those obtained for random sequences of the amino acids of the same peptide. We have assumed that all codons for each amino acid are equally likely. To keep the computing effort reasonably small, we have limited our attention to small peptides (table 1), the amino acid sequences for which have been published[2-4]. Furthermore, we have avoided analysis of the possible simultaneous presence of several bending positions or presence of stems with stretches of noncontiguous pairing and we have confined our attention to extensive runs of contiguous pairs. Our findings have made it unlikely that this is a serious limitation. For each peptide the longest run of contiguous pairs was compared with runs obtained from random sequences of those same amino acids.

Results. Except for ranatensin, no more than a hundred peptides generated by random arrangement of the constituent amino acids was needed to yield at least one run of paired bases as long as or longer than the longest obtained for the natural peptide (table 1, column 5). Furthermore, the fraction of random peptide sequences which gave rise to a run equal to or longer than the maximum obtained for the natural peptide is often substantial, and from them one can obtain the 95% confidence intervals for binomial distribution (table 1, footnote). Thus, for example, for ranatensin there is a 95% chance that the 'true' percentage lies between 0–4%.

Table 1. Data for runs of base pairs in hypothetical m-RNA's of natural peptides (columns 1-3) and random amino acid sequences of the same natural peptides (columns 4 and 5)

	No. of amino acids	No. of random sequences studied	Maximum No. of bases paired in a run		No. of random sequences with run of pairs $>N_p(=N_p)$
			Natural peptide (N_p)	Random peptide sequence	
	(1)	(2)	(3)	(4)	(5)
Angiotensinogen (horse)**	10	30	20	20	0 (3)
Bradykinin (ox)	9	10	12	20	8 (1)
Met-Lys bradykinin (ox)	11	100	16	26	33 (29)
Bradykinin potentiating peptide B (snake)	11	10	8	16	10 (−)
Eledoisin (octopus)	11	10	14	18	2 (7)
Gastrin (pig)	17	20	12	20	8 (6)
Glumitocin (fish)	9	10	10	22	8 (2)
Growth hormone releasing hormone (pig)	10	10	14	16	1 (3)
Isotocin (fish)	9	100	16	22	10 (18)
Kallidin (ox)	10	10	12	24	9 (1)
Kininogen (human)	11	10	16	20	5 (4)
Luteinising hormone releasing hormone (pig)	10	100	18	22	9 (18)
Mast cell degranulating peptide (bee)**	22	30	22	26	1 (3)
α-Melanocyte stimulating hormone (dogfish)	11	100	10	20	89 (11)
β-Melanocyte stimulating hormone (ox)**	18	10	26	28	1 (0)
Mesotocin (frog)	9	10	10	24	8 (1)
Motilin (pig)	22	10	18	26	6 (3)
Phyllomedusin (amphibian)	10	20	18	20	3 (2)
Physalaemin (amphibian)	11	10	14	11	2 (1)
Ranatensin (frog)**	11	107	22	20	0 (0)
Somatostatin (sheep)**	14	100	26	30	2 (0)
Substance P (ox)**	11	40	18	18	0 (2)
Vasopressin (ox)	9	20	10	14	10 (9)
Vasotocin (ox)	9	10	10	16	5 (4)
Thyrotrophin releasing hormone (pig)	3	6*	4	8	2 (3)
Leu-enkephalin (pig)	5	60*	6	10	22 (28)
Met-enkephalin (pig)	5	60*	6	8	8 (25)

* All possible random sequences; ** 95% confidence intervals (per cent) for the percentage of random sequences with pairing equal to or greater than the natural sequence are respectively: 2-27%, 4-31%, 0-45%, 0-4%, 0-7% and 0-18%.

Table 2. Frequency table of the numbers of natural sequences and random amino acid sequences subdivided into class intervals according to the maximum fraction length of run. The latter is the number of bases involved in the longest run of contiguous paired bases expressed as a fraction of the total bases in the sequence. The natural sequences are those of the peptides in table 1; the random sequences are for the 6 peptides in table 1 where 100 random sequences of the constituent amino acids were considered for each peptide.

Maximum fractional run of base pairs	<0.4	0.40–0.49	>0.49
Natural sequences	9	11	7
Random sequences	146	291	170

$\chi^2 = 1.25$, d.f. 2, $p > 0.5$.

The maximum length of run (columns 3 and 4) may be expressed as a fraction of the number of bases in the respective hypothetical m-RNA's. For the 6 peptides for which 100 random sequences were generated it was found that the numerical distribution of peptides among class intervals of maximum fractional run lengths is not different (χ^2-test) for natural peptides and random sequences (table 2).

Discussion. Thus we have found that, contrary to the conclusions of Polya (1975) concerning somatostatin and our preliminary conclusion about angiotensinogen, long runs of paired bases can readily occur. Furthermore, the distribution of run lengths for a group of natural peptides is not significantly different from that for peptide sequences obtained by the random arrangement of the amino acids making up those natural peptides. There is no indication that peptides at the extremes of this distribution (e.g. ranatensin and somatostatin) comprise subsets not present in the distribution obtained for the random arrangements of amino acids. Therefore, a requirement for extensive base pairing in m-RNA to give it stability would not impose serious restrictions on the amino acid sequence. Thus it is not possible from a study of the amino acid sequence of peptides to draw conclusions about the importance of pairing in the m-RNA's.

2 points need further comment. First, we have not considered the possibility of either non-Watson-Crick pairing or the simultaneous presence of 2 or more runs of base pairings for any particular base sequence. Since both possibilities would increase the degree of pairing our disregarding them cannot weaken our conclusion. Secondly, we have assumed that all codons for each amino acid are equally likely, but recent evidence[5] for one m-RNA contradicts this assumption. Although the failure of this assumption would weaken our conclusion, its effect should be small in view of the high probability of pairing shown for the random sequences of all but one peptide.

1 G.M. Polya, Nature *255*, 652 (1975).
2 M.O. Dayhoff, Atlas of Protein Sequences and Structure, vol. 4. National Biomedical Research Foundation, Silver Spring, Maryland 1972.
3 L.R. Croft, Handbook of Protein Sequences, and suppl. A. Joynson-Bruvvers, Oxford 1973 and 1974.
4 J. Hughes, T.W. Smith, H.W. Kosterlitz, L.A. Fothergill, B.A. Morgan and H.R. Morris, Nature *258*, 577 (1975).
5 L. McReynolds, B.W. O'Malley, A.D. Nisbet, J.E. Fothergill, D. Givol, S. Fields, M. Robertson and G.C. Brownlee, Nature *273*, 723 (1978).