Binding of peptides that are specified by complementary RNAs

J. Edwin BLALOCK and Kenneth L. BOST

Department of Microbiology, University of Texas Medical Branch, Galveston, TX 77550, U.S.A.

An interesting pattern in the genetic code was reported previously [Blalock & Smith (1984) *Biochem. Biophys. Res. Commun.* **121**, 203–207]. In the 5'-to-3' direction, codons for hydrophilic and hydrophobic amino acids are generally complemented by codons for hydrophobic and hydrophilic amino acids respectively. The average tendency of codons for 'uncharged' (slightly hydrophilic) amino acids was to be complemented by codons for 'uncharged' amino acids. We now show that the same pattern results when the complementary codon is read in the 3'-to-5' direction. This pattern is further shown to result in the interaction of peptides specified by complementary RNAs regardless of whether the amino acids are assigned in the 5'-to-3' or the 3'-to-5' direction. Here we demonstrate that peptides specified by complementary RNAs bind to each other with specificity and high affinity.

INTRODUCTION

We have previously reported that, when read in the 5'-to-3' direction, codons for hydrophilic and hydrophobic amino acids are generally complemented by codons for hydrophobic and hydrophilic amino acids respectively. The average tendency of codons for 'uncharged' (slightly hydrophilic) amino acids was to be complemented by codons for 'uncharged' amino acids (Blalock & Smith, 1984). In the present paper we show that the same pattern results when the complementary codon is read in the 3'-to-5' direction. We previously theorized that a result of the original 5'-to-3' relationship is that complementary DNAs, when transcribed in this direction and in the same reading frame, will code for peptides or proteins that interact. This theory was tested and supported by the observation of specific and high-affinity binding of naturally occurring peptides (ACTH and γ -endorphin) to synthetically derived counterparts that were specified by the complementary RNA sequences for ACTH mRNA and γ -endorphin mRNA respectively (Bost *et al.*, 1985a). The present observation of hydropathic anti-complementarity of amino acids for complementary codons read in the 3'-to-5' direction suggests that complementary RNAs should code for interacting peptides regardless of whether amino acids are assigned in the 5'-to-3' or the 3'-to-5' direction. Experiments reported in the present paper support this idea and detail the pattern of hydropathic anti-complementarity of amino acids for complementary codons read in the 3'-to-5' direction.

MATERIALS AND METHODS

Peptides

Peptides were synthesized according to either the 5'-to-3' or 3'-to-5' sequences complementary to the mRNA of bovine ACTH-(1–24)-peptide (Nakanishi *et al.*, 1979) by Peninsula Laboratories (San Carlos, CA, U.S.A.). The final products were estimated to be 91% intact peptide and 9% fragments as analysed by h.p.l.c. and compositional amino acid analysis. Synthetic ACTH

was obtained from Organon Laboratories (West Orange, NJ, U.S.A.) and 125 I-labelled ACTH (200 μ Ci/ μ g) from New England Nuclear (Boston, MA, U.S.A.).

Binding assay

A solid-phase binding assay was used to determine the ability of both the 5'-to-3' and the 3'-to-5'-complementary-mRNA-strand peptides to bind 125 I-ACTH. The peptides were diluted in phosphate-buffered saline (140 mM-NaCl/3 mM-KCl/0.1% NaN_3 /20 mM-phosphate buffer, pH 7.4) to 2 mg/ml, and 0.2 ml/well was placed into poly(vinyl chloride) micro-titre plates (Becton and Dickinson, Oxnard, CA, U.S.A.). After 18 h at 4 °C, unbound peptide was removed by washing three times with phosphate-buffered saline containing 0.1% bovine serum albumin (Sigma Chemical Co., St. Louis, MO, U.S.A.). 125 I-ACTH was diluted to various concentrations with phosphate-buffered saline containing 0.1% bovine serum albumin and incubated in peptide-coated wells for 60 min at 4 °C. To some wells, various concentrations of soluble peptides were added in addition to the 125 I-ACTH in order to demonstrate the specificity of binding. After incubation, unbound radiolabel was washed out with phosphate-buffered saline containing 0.1% bovine serum albumin, and the radioactivity of individual wells was measured by γ -radiation counting. The amount of specifically bound radioactivity was determined by blocking 125 I-ACTH binding with unlabelled ACTH. Additional controls included coating wells with insulin or bovine serum albumin (2 mg/ml) before the addition of the 125 I-ACTH.

RESULTS

3'-to-5'-complementary-mRNA-strand amino acids

As previously described (Blalock & Smith, 1984), the anti-complementarity of amino acids based on the genetic code is most obvious when presented in terms of the hydropathic character of amino acids. By employing the hydropathic scoring system of Kyte & Doolittle (1982),

Abbreviation used: ACTH, corticotropin.

Table 1. Hydropathic scores for the amino acids whose codons are complementary to those of hydrophobic amino acids

Coding strand			Non-coding strand				
Codon	Amino acid	Hydropathic score	Codon	Amino acid	Hydropathic score		Average change
					Individual	Average	
AUA	Isoleucine	+4.5	UAU	Tyrosine	−1.3	−1.3	−5.8
GUU	Valine	+4.2	CAA	Glutamine	−3.5	−3.4	−7.6
GUC	Valine		CAG	Glutamine	−3.5		
GUG	Valine		CAC	Histidine	−3.2		
GUA	Valine		CAU	Histidine	−3.2		
UUA	Leucine	+3.7	AAU	Asparagine	−3.5	−3.5	−7.2
UUG	Leucine		AAC	Asparagine	−3.5		
CUU	Leucine		GAA	Glutamic acid	−3.5		
CUC	Leucine		GAG	Glutamic acid	−3.5		
CUA	Leucine		GAU	Aspartic acid	−3.5		
CUG	Leucine		GAC	Aspartic acid	−3.5		
UUU	Phenylalanine	+2.7	AAA	Lysine	−3.9	−3.9	−6.6
UUC	Phenylalanine		AAG	Lysine	−3.9		
UGU	Cysteine	+2.5	ACA	Threonine	−0.7	−0.7	−3.2
UGC	Cysteine		ACG	Threonine	−0.7		
AUG	Methionine	+1.9	UAC	Tyrosine	−1.3	−1.3	−3.2
GCU	Alanine	+1.8	CGA	Arginine	−4.5	−4.5	−6.3
GCC	Alanine		CGG	Arginine	−4.5		
GCA	Alanine		CGU	Arginine	−4.5		
GCG	Alanine		CGC	Arginine	−4.5		

Table 2. Hydropathic scores for the amino acids whose codons are complementary to those of hydrophilic amino acids

Coding strand			Non-coding strand				
Codon	Amino acid	Hydropathic score	Codon	Amino acid	Hydropathic score		Average change
					Individual	Average	
CGU	Arginine	−4.5	GCA	Alanine	+1.8	+0.9	+5.4
CGC	Arginine		GCG	Alanine	+1.8		
CGA	Arginine		GCU	Alanine	+1.8		
CGG	Arginine		GCC	Alanine	+1.8		
AGA	Arginine		UCU	Serine	−0.9		
AGG	Arginine		UCC	Serine	−0.9		
AAA	Lysine	−3.9	UUU	Phenylalanine	+2.7	+2.7	+6.6
AAG	Lysine		UUC	Phenylalanine	+2.7		
AAU	Asparagine	−3.5	UUA	Leucine	+3.7	+3.7	+7.2
AAC	Asparagine		UUG	Leucine	+3.7		
GAU	Aspartic acid	−3.5	CUA	Leucine	+3.7	+3.7	+7.2
GAC	Aspartic acid		CUG	Leucine	+3.7		
CAA	Glutamine	−3.5	GUU	Valine	+4.2	+4.2	+7.7
CAG	Glutamine		GUC	Valine	+4.2		
GAC	Glutamic acid	−3.5	CUG	Leucine	+3.7	+3.7	+7.2
GAA	Glutamic acid		CUU	Leucine	+3.7		
CAG	Histidine	−3.2	GUG	Valine	+4.2	+4.2	+7.4
CAU	Histidine		GUA	Valine	+4.2		

amino acids were assigned a numerical value based on their relative hydrophobic (+ side) or hydrophilic (− side) nature. The hydropathic scale extends from +4.5 (isoleucine) to −4.5 (arginine) and allows for the grouping of amino acids as hydrophobic (+4.5 to +1.8; Table 1), hydrophilic (−4.5 to −3.2; Table 2) and

slightly hydrophilic (−0.4 to −1.6; Table 3). Table 1 presents the codons for the hydrophobic amino acids and their hydropathic scores followed by the complementary codons (non-coding strand) read in the 3′-to-5′ direction and their corresponding amino acids. Interestingly, of the possible 20 complementary codons, there were none that

Table 3. Hydropathic scores for the amino acids whose codons are complementary to those of slightly hydrophilic amino acids

Coding strand			Non-coding strand				
Codon	Amino acid	Hydropathic score	Codon	Amino acid	Hydropathic score		
					Individual	Average	Average change
GGU	Glycine	-0.4	CCA	Proline	-1.6	-1.6	-1.2
GGC	Glycine		CCG	Proline	-1.6		
GGA	Glycine		CCU	Proline	-1.6		
GGG	Glycine		CCC	Proline	-1.6		
ACC	Threonine	-0.7	UGG	Tryptophan	-0.9	+1.4	+2.1
ACG	Threonine		UGC	Cysteine	+2.5		
ACA	Threonine		UGU	Cysteine	+2.5		
UGG	Tryptophan	-0.9	ACC	Threonine	-0.7	-0.7	+0.2
UCU	Serine		AGA	Arginine	-4.5		
UCC	Serine		AGG	Arginine	-4.5		
UCA	Serine	-0.9	AGU	Serine	-0.9	-2.1	-1.2
UCG	Serine		AGC	Serine	-0.9		
AGU	Serine		UCA	Serine	-0.9		
AGC	Serine		UCG	Serine	-0.9		
UAU	Tyrosine	-1.3	AUA	Isoleucine	+4.5	+3.2	+4.5
UAC	Tyrosine		AUG	Methionine	+1.9		
CCU	Proline	-1.6	GGA	Glycine	-0.4	-0.4	+1.2
CCC	Proline		GGG	Glycine	-0.4		
CCA	Proline		GGU	Glycine	-0.4		
CCG	Proline		GGC	Glycine	-0.4

Fig. 1. Relationship of the hydropathic scores of the coding-strand amino acids and the hydropathic scores of the corresponding non-coding-strand amino acids read in the 3'-to-5' direction

code for hydrophobic amino acids; 16 were hydrophilic and four were slightly hydrophilic.

An opposite pattern was once again observed for the hydrophilic amino acids (Table 2). Of the possible 18 complementary codons, none coded for strongly hydrophilic amino acids; 16 were for hydrophobic amino acids and two were for slightly hydrophilic amino acids, and all of these were complementary codons for a single amino acid (arginine).

The pattern for the slightly hydrophilic amino acids once again differed from that of the two previous groups

(Table 3). Of the possible 20 complementary codons, 14 code for slightly hydrophilic and two for strongly hydrophilic amino acids. The two complementary codons for strongly hydrophilic amino acids are roughly balanced by four codons for hydrophobic amino acids. Thus the net effect is for little change in the average hydropathic character of the non-coding-strand amino acids. Tyrosine was again the single amino acid in this group that showed a major average change in hydropathic scores of its non-coding-strand amino acids. Fig. 1 shows a plot of the hydropathic scores of the coding-strand amino acids versus the hydropathic scores of the non-coding-strand amino acids. A linear-regression analysis of the data results in a correlation coefficient of -0.77 . Very similar results were obtained when the hydropathic scoring system of Hopp & Woods (1981) was employed (results not shown). Thus, as was the case for the 5'-to-3' direction, in the 3'-to-5' direction the non-coding-strand hydropathic scores are inversely related to the coding strand, and this relationship is not random.

Specific binding of peptides encoded by complementary RNA read in the 5'-to-3' or the 3'-to-5' direction

To determine whether the above-described pattern would result in an interacting peptide, peptides corresponding to the complementary RNA sequence of ACTH-(1-24)-peptide mRNA were synthesized according to the complementary sequence being read either 5'-to-3' or 3'-to-5'. These peptides have the respective sequences $\text{NH}_2\text{-Gly-Val-His-Leu-His-Arg-Ala-Pro-Leu-Leu-Ala-His-Arg-Leu-Ala-Pro-Ala-Glu-Val-Phe-His-Gly-Val-Arg-CO}_2\text{H}$ and $\text{NH}_2\text{-Arg-Met-Arg-Tyr-Leu-Val-Lys-Ala-Thr-Pro-Phe-Gly-His-Pro-Phe-Phe-Ala-Ala-Gly-His-Phe-His-Met-Gly-CO}_2\text{H}$ and were assayed for the

Fig. 2. Binding of ^{125}I -ACTH to the 5'-to-3'- or 3'-to-5'-complementary-mRNA-strand peptides

Dilutions of ^{125}I -ACTH were incubated in wells coated with 5'-to-3'-complementary-mRNA-strand peptide (●) or 3'-to-5'-complementary-mRNA-strand peptide (○), and the amount of specifically bound radiolabel was determined. Unlabelled ACTH blocked greater than 84% of the ^{125}I -ACTH binding.

Fig. 3. Ability of soluble 5'-to-3'- or 3'-to-5'-complementary-mRNA-strand peptides to block the solid-phase binding of ^{125}I -ACTH

Wells were coated with 5'-to-3'-complementary-mRNA-strand peptide (●) or 3'-to-5'-complementary-mRNA-strand peptide (○), and then ^{125}I -ACTH was incubated in wells in the presence of various concentrations of the appropriate soluble peptide. In the presence of soluble peptide, greater than 90% of the ^{125}I -ACTH binding was blockable. In addition, when insulin (▲) or bovine serum albumin (■) was used to coat wells, less than 5% of the ^{125}I -ACTH binding was non-specific.

ability to bind ACTH. Fig. 2 shows that both peptides were capable of binding ^{125}I -ACTH. Greater than 84% of the binding was blocked by non-labelled ACTH, and half-maximal binding occurred at 0.3 nM-ACTH. These curves were similar to what one would expect for a theoretical curve for dissociation constant of 0.3 nM. Fig. 3 demonstrates the specificity of the peptide- ^{125}I -ACTH interaction, since the addition of various

concentrations of soluble peptide blocked ^{125}I -ACTH binding in a dose-dependent manner. When compared with the ability of the synthetic peptides to bind ^{125}I -ACTH, less than 5% of the total ^{125}I -ACTH bound non-specifically to insulin- or bovine serum albumin-coated wells (Fig. 3). Thus the binding of ACTH by both peptides was specific and of high affinity. Furthermore, there was essentially no difference between the abilities of the 3'-to-5'- and the 5'-to-3'-complementary-mRNA-strand peptides to bind ACTH or to block ^{125}I -ACTH binding.

DISCUSSION

An interesting pattern in the genetic code that allows for hydrophathic anti-complementary of amino acids was reported previously (Blalock & Smith, 1984). In the present paper we have shown that an almost identical pattern results when the complementary codon is read in the 3'-to-5' rather than the 5'-to-3' direction. Since, regardless of the reading direction, the second base of the codon never changes, it apparently is responsible for the hydrophathic anti-complementarity of amino acids that are specified by complementary codons. This seems to result largely from the fact that the preponderance (six of seven) of hydrophilic amino acids have A as their second base codon whereas the complementary base, U, is the second for most (five of seven) hydrophobic amino acids. One of the two exceptions to the above in the hydrophobic group (alanine) does not matter as it has a second base, C, and the second base for the single exception in the hydrophilic group (arginine) has a second base, G. Hence there is a virtually perfect interchange of hydrophobic and hydrophilic amino acids whether the complementary codon is read 5'-to-3' or 3'-to-5'. With the exception of tyrosine, the second base of the other five uncharged (slightly hydrophilic) amino acids is either a G or C. Hence the codons for this group will usually result in a similar type of amino acid regardless of the direction in which the complementary codon is read.

A case can be made that the genetic code arose during evolution as a result of the chemical similarity of anticodonic bases and their respective amino acids (Weber & Lacey, 1978; Jungck, 1978). Perhaps this similarity resulted in the pattern that we have observed. It is tempting to speculate that a functional and evolutionary advantage to this resided in the fact that the second base of codons for similar amino acids is the same. Perhaps, before the advent of the directionality of nucleic acid reading, an amino acid from the same hydrophathic group would be present and thus the resulting peptides or proteins would be grossly similar in conformation, whether nucleic acids were read 5'-to-3' or 3'-to-5'. More importantly, this hypothesis predicts that peptide products of complementary nucleic acid strands would bind each other regardless of the direction of amino acid assignment. To test this hypothesis, peptides encoded by the mRNA complementary to the mRNA for ACTH were synthesized in both the 5'-to-3' and the 3'-to-5' directions. These peptides bound ^{125}I -ACTH with specificity and high affinity (approx. $3 \times 10^9 \text{ M}^{-1}$). The probability of randomly selecting two peptides that bind ACTH with such high affinities seems remote. Therefore these data provide support for a simple and foolproof method for the evolution of interacting peptide or protein pairs. Additional support for such a notion seems evident

in our observation of regions of complementarity between mRNAs for ligands (epidermal growth factor, transferrin and interleukin) and their receptors when such RNAs are aligned from their 5' ends (Bost *et al.*, 1985*b*).

This research was supported by a grant from Triton Biosciences and by a McLaughlin Postdoctoral Fellowship to K. L. B.

REFERENCES

- Blalock, J. E. & Smith, E. M. (1984) *Biochem. Biophys. Res. Commun.* **121**, 203–207
- Bost, K. L., Smith, E. M. & Blalock, J. E. (1985*a*) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1372–1375
- Bost, K. L., Smith, E. M. & Blalock, J. E. (1985*b*) *Biochem. Biophys. Res. Commun.* **128**, 1373–1380
- Hopp, T. P. & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3824–3828
- Jungck, J. R. (1978) *J. Mol. Evol.* **11**, 211–224
- Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
- Nakanishi, S., Inoue, A., Kita, T., Nakamura, M., Chang, A. C. Y., Cohen, S. N. & Numa, S. (1979) *Nature (London)* **278**, 423–427
- Weber, A. L. & Lacey, J. C. (1978) *J. Mol. Evol.* **11**, 199–210

Received 9 April 1985/31 October 1985; accepted 13 November 1985