Vol. 148, No. 1, 1987 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
October 14, 1987 Pages 456-462

HYDROGEN BONDING BETWEEN CYTOSINE AND PEPTIDES OF THREONINE OR SERINE: IS IT RELEVANT TO THE ORIGIN OF THE GENETIC CODE?

Chien-Hua Niu[1], Kyou-hoon Han[1], Herman J. C. Yeh[2], and Simon Black[3]

[1]Laboratory of Experimental Carcinogenesis, Division of Cancer Etiology, Building 37, Room 3C03, National Cancer Institute, [2]Laboratory of Analytical Chemistry, Building 8, Room B2A-22, and [3]Laboratory of Biochemical Pharmacology, Building 8, Room 215, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

Received July 13, 1987

^{13}C, ^{15}N, and ^{1}H nuclear magnetic resonance measurements indicate that chloroform-soluble threonine-containing tripeptide derivatives, such as t-Boc-Thr-Gly-Gly-OBz, form three strong hydrogen bonds to the cytosine moiety of 2',3'-O-isopropylidene-5'-O-t-butyldimethylsilylcytidine. The C=O and NH of the central peptide residue plus the OH of the threonine side chain appear to form bonds to the $N(4')H_2$, N(3), and C(2)=O, respectively, of the pyrimidine. An association constant calculated from the cytidine $^{15}N(4')$ nuclear magnetic resonance response to added peptide is four times larger than the corresponding cytosine-guanine constant. It is suggested that cytosine-peptide bonding was part of the primitive genetic coding mechanism early in evolution and accounts for the origin of the cytosine-centered codons for the hydroxy amino acids, serine and threonine, in the present code.

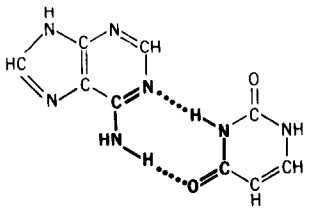
In organisms existing at the present time, the recognition of an amino acid by a codon is effected indirectly through a molecule of tRNA, but recognition early in evolution may have occurred through direct amino acid-template interactions (1, 2). The present role of uracil as the central base of codons exclusively for hydrophobic amino acids and of cytosine as the central base of eight of the ten codons for the hydroxy amino acids, serine and threonine (Table 1), may be vestiges of such an earlier mechanism. A hypothetical primitive template was proposed (3, 4) in which uracil had side chains that recognized specific hydrophobic amino acid side chains coordinately with formation of two hydrogen bonds to the peptide backbone (Fig. 1, lower left). Cytosine was also proposed (5) to have formed two hydrogen bonds to the peptide backbone plus a third one to the hydroxyl group of a serine or threonine residue (Fig. 1, lower right). The present experiments demonstrate the feasibility of the cytosine-peptide hydrogen bonding. Uracil-peptide bonding was previously reported (6).

Simple monomeric nucleotides form hydrogen bonds to each other in water only in extremely high concentration and possibly only as members of hydrophobically-formed stacks rather than as unassociated molecules (7). In organic solvents, however, these compounds bond readily, and Watson-Crick

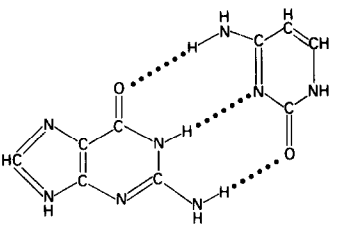
0006-291X/87 $1.50

Table 1. Genetic code arrangement of amino acids having hydrocarbon and hydroxy side chains. The symbols for the hydroxy amino acids, serine and threonine, are enclosed in frames. It is suggested that this arrangement reflects a left to right evolution of the code, beginning with hydrophobic and hydroxy amino acids only and operating through a primitive mechanism. Amino acids with adenine- and guanine-centered codons may have entered the picture through the very different tRNA mechanism, which would explain the presence of a very different second set of serine codons. A possible derivation of primitive proline and alanine codons from those for the similarly shaped molecules of threonine and serine is suggested in the Discussion.

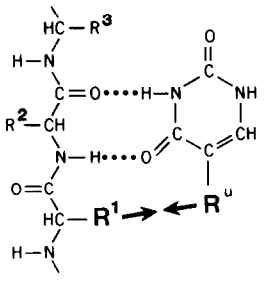
	U	C	A	G	
U	phe	ser			U
	phe	ser			C
	leu	ser			A
	leu	ser			G
C	leu	pro			U
	leu	pro			C
	leu	pro			A
	leu	pro			G
A	ile	thr		ser	U
	ile	thr		ser	C
	ile	thr			A
		thr			G
G	val	ala			U
	val	ala			C
	val	ala			A
	val	ala			G



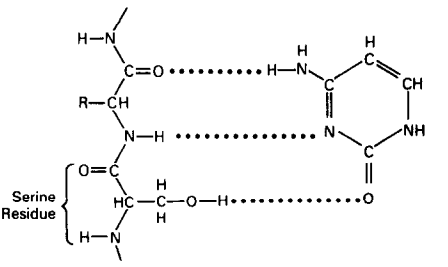
Adenine •••• Uracil



Guanine •••• Cytosine



Peptide ••• Uracil-R



Peptide•••••••••••• Cytosine

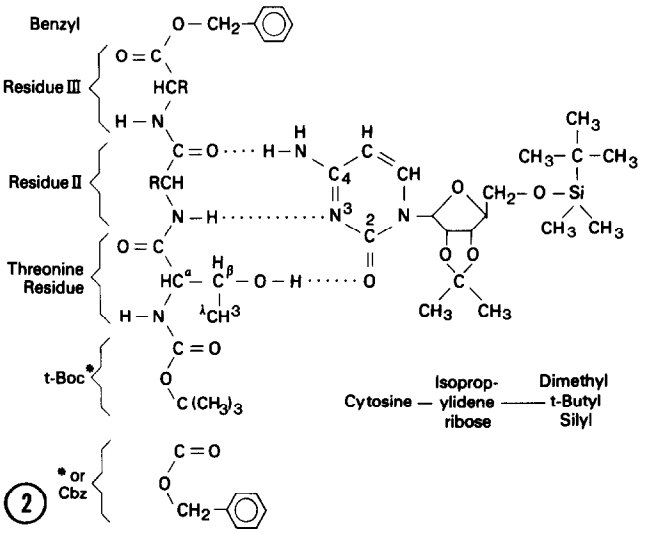
Fig. 1. Watson-Crick hydrogen-bonding between nucleotide bases (adenine-uracil and guanine-cytosine) and postulated analogous base-peptide binding in a hypothetic1 primitive genetic coding process.

pairing is especially strong (8-11). Hydrogen bonding in organic solvents may therefore serve as a provisional measure of the potential of monomeric compounds, when linked in appropriate polymers, to form hydrogen bonds in water.

MATERIALS AND METHODS See figure legends.

RESULTS

The Peptide C=O(II) - Cytosine $N(4')H_2$ Bond. On gradual addition of t-Boc-Thr-Gly-Gly-OBz to the cytosine derivative shown in Fig. 2, labeled in the N(4') position with ^{15}N, the ^{15}N resonance shifted downfield to reach a plateau at three peptide equivalents per unit of base (Fig. 3). Association of this shift with hydrogen bonding was confirmed by substituting tri(t-butyldimethylsilyl)guanosine for the peptide to form a Watson-Crick pair. The C=O(II) of the peptide (Fig. 2) appears to be the sole receptor, for only its ^{13}C resonance among the eligible groups moved downfield on addition of the cytosine derivative (Table 2). At -20°C in chloroform the $N(4')H_2$ proton resonances were observed at 5.08 and 7.31 ppm without peptide and at 5.75 and 8.4 ppm in the presence of one equivalent of peptide.



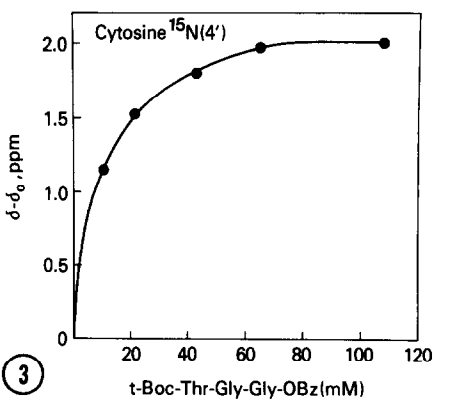


Fig. 2. Formulae of the major chloroform-soluble substances used in the experiments described here. Peptide-cytosine hydrogen bonds inferred from the experiments are indicated by dotted lines.

Fig. 3. Change in the $^{15}N(4')$ chemical shift on addition of t-Boc-L-Thr-Gly-Gly-OBz to a chloroform-soluble $^{15}N(4')$ cytosine derivative. t-Boc-L-Thr-Gly-Gly-OBz was synthesized by the DCC-HOBT [1-ethyl-3-(3-dimethylaminopropyl)carbodiimide and 1-hydroxybenzotriazole] method and the $^{15}N(4')$ cytosine derivative (Fig. 2) as previously described (18). The cytosine derivative concentration was 21 mM in deuterated chloroform. [^{15}N]NMR spectra were obtained at 27.36 MHz with an NIH-modified spectrometer with a Brucker superconducting magnet equipped with a Nicolet Model 1180 computer. A 10-mm probe was used at an operating temperature of 20°C.

Table 2. ^{13}C NMR chemical shifts for free t-Boc-L-Thr-Gly-Gly-OBz and its complex with 2',3'-O-isopropylidene-5'-O-t-butyldimethylsilyl cytidine. The peptide concentration was kept at 52.28 mM for measurements with both the free peptide and the complex formed on addition of one equivalent of the cytidine derivative. The solvent was [^{2}H]chloroform, and the temperature was 20°C. Chemical shifts are given in parts per million downfield from TMS. The chemical shift assignments are based on the undecoupled spectrum. ^{13}C Carbonyl group assignment was made with two different peptides having 20% ^{13}C enrichment in C=O(II) and C=O(III), respectively. The two glycine α-carbon resonances were not specifically identified.

Peptide carbon	Free peptide	Peptide-cytidine complex	Difference
C=O(I)	172.25	172.16	0.09
C=O(II)	169.59	170.03	- 0.44
C=O(III)	170.08	170.03	0.05
C=O (t-Boc)	156.40	156.29	0.11
Thr α	59.44	59.54	- 0.10
Thr β	67.39	67.66	- 0.27
Thr γ	18.97	18.71	0.26
Gly α	43.01	43.05	- 0.04
Gly α	41.28	41.15	0.13

The Peptide OH - Cytosine C(2)=O Bond. Addition of t-Boc-Thr-Gly-Gly-OBz to the cytosine derivative causes a downfield shift in its $^{13}C(2)=O$ resonance (Fig. 4) comparable to the ^{13}C shifts of the carbonyl groups of uracil when hydrogen-bonded to adenine in chloroform (11). Several lines of evidence indicate that the threonine OH is the major donor in this case: (a) In the presence of one equivalent of the cytosine derivative, the OH-bearing β-carbon of the threonine residue of t-Boc-Thr-Gly-Gly-OBz shows a substantial downfield shift in its ^{13}C resonance and the γ-carbon shows a substantial upfield shift (Table 2), both supportive of participation of the OH in hydrogen bonding. (b) Experiments with t-Boc-Gly-Gly-Gly-OBz showed that three of its

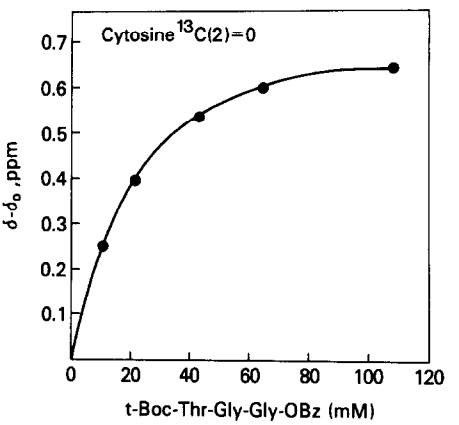
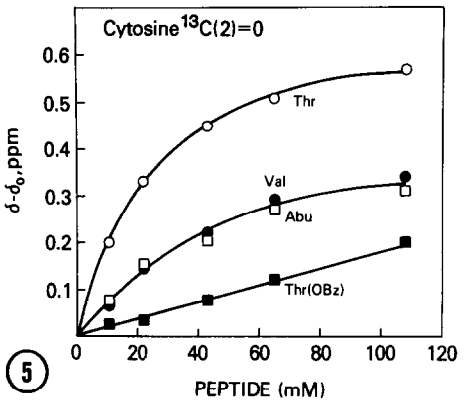


Fig. 4. Change in the $^{13}C(2)=O$ chemical shift on addition of t-Boc-L-Thr-Gly-Gly-OBz to a chloroform-soluble cytosine derivative. The ^{13}C spectra were obtained at 67.89 MHz with other conditions as described in the legend to Fig. 3.



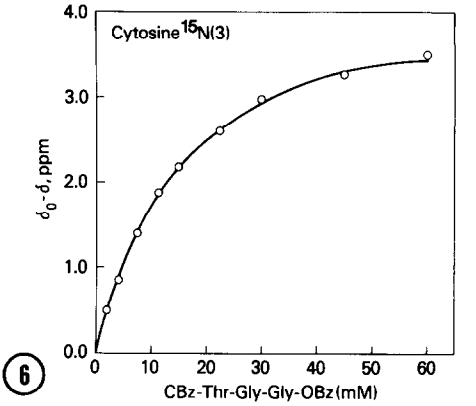


Fig. 5. Change in the ^{13}C(2)=0 chemical shift of a chloroform-soluble cytosine derivative on addition of t-Boc-L-Thr-L-Ala-L-Ala-OBz (o), t-Boc-L-Abu-L-Ala-L-Ala-OBz (□), t-Boc-L-Val-L-Ala-L-Ala-OBz (●), and t-Boc-L-Thr(OBz)-L-Ala-L-Ala-OBz (■). Abu = α-aminobutyric acid. The peptide derivatives were synthesized as indicated for t-Boc-Thr-Gly-Gly-OBz in the legend to Fig. 3. Other conditions were as described in the legends to Figs. 3 and 4.

Fig. 6. Change in the ^{15}N(3) chemical shift on addition of Cbz-L-Thr-Gly-Gly-OBz to a chloroform-soluble (1,3-$^{15}N_2$)cytidine derivative. The synthesis of the ^{15}N-labeled cytosine derivative was previously described (18). Its concentration in $CDCl_3$ was 21 mM. The experiment was performed with a JEOL FX-100 NMR spectrometer at 10.14 MHz. A 10-mm probe was used at 20°C.

four C=O groups form hydrogen bonds with cytosine (5). In contrast, the presence of threonine as residue I of an otherwise identical peptide eliminates binding to C=O groups I and III, which are not involved in the formulation of Fig. 2 (Table 2). Thus the threonine OH adds sufficient binding strength to preserve the C=O(II) - N(4')H_2 bond of Fig. 2, excluding certain competing structures that form in its absence. (c) An apparent association constant for cytosine - t-Boc-Thr-Gly-Gly-OBz calculated (12) from the data of Fig. 3 was 285 M^{-1}, significantly larger than the value of 136 M^{-1} found in a comparable experiment for cytosine - t-Boc-Gly-Gly-Gly-OBz, confirming the conclusion that the OH group is a major contributor to the binding strength between threonine peptides and cytosine. For comparison, an association constant of 67 M^{-1} was found for the Watson-Crick pairing of the cytidine derivative with tri(t-butyldimethylsilyl)guanosine in chloroform. (d) t-Boc-Thr-Ala-Ala-OBz produces a greater shift in the cytosine ^{13}C(2)=O resonance (Fig. 5) than do its analogs in which the OH is (i) absent (t-Boc-Abu-Ala-Ala-OBz), (ii) replaced with CH_3 (t-Boc-Val-Ala-Ala-OBz), or (iii) blocked as the benzyl ether [t-Boc-Thr(OBz)-Ala-Ala-OBz]. In the absence of the threonine or serine OH group, the binding to the cytosine C(2)=O appears to be through an imide proton (5).

The Peptide NH(II) - Cytosine N(3) Bond. Addition of t-Boc-Thr-Gly-Gly-OBz to a (1,3-$^{15}N_2$)cytosine derivative produces a large upfield shift in the ^{15}N(3) resonance (Fig. 6), implicating this nitrogen atom as a hydrogen

bond receptor (13). The $^{15}N(1)$ resonance does not respond to peptide addition. In comparable experiments with unlabeled cytosine and derivatives of triglycine suitably enriched with deuterium to accentuate the imide proton resonances, the NH(II) was the most prominent peptide donor in hydrogen bonding, supporting in that case a formulation analogous to the one in Fig. 2 (5).

Serine Versus Threonine. In an experiment comparable to that of Table 2, the ^{13}C resonance of the β-carbon of the serine residue in t-Boc-Ser-Gly-Gly-OBz shifted downfield 0.13 ppm in the presence of one equivalent of the cytosine derivative, whereas the resonance of the corresponding threonine carbon of t-Boc-Thr-Gly-Gly-OBz shifted 0.27 ppm. Apparently the electron-donating methyl group of threonine potentiates the tendency of the OH group to form a hydrogen bond. The upfield shift of the ^{13}C resonance of the methyl (γ) carbon supports this view (Table 2).

DISCUSSION

In a molecular model, three hydrogen bonds form readily between cytosine and a serine- or threonine-containing peptide according to the formulations of Figs. 1 and 2 to produce an intimately fitting complementary structure. It is improbable that another hypothetical structure would so completely accommodate all of the chemical shift changes reported here.[1]

A larger model shows that a closely-fitting helical nucleic acid-like structure can be built around a polypeptide that is hydrogen-bonded to every third base (3). Each primitive cytosine-centered codon may have been completed by side chains extending from the two bases flanking cytosine, and the coding pockets thus formed for threonine and serine might then have evolved to accommodate the similar forms of proline and alanine, accounting for their presence in the cytosine column of Table 1. The original templates, precisely fitted to specific peptide structures, would be very different from their

[1]Several efforts were made to obtain further support for the proposed hydrogen bonding network through nuclear Overhauser enhancement (NOE) spectroscopy. Conventional ^1H homonuclear two-dimensional (2D) NOE (NOESY) (14, 15) and 2D spin-locked NOE (ROESY) (16, 17) experiments, both at 500 MHz and at -20°C and 25°C, respectively, were employed to measure NOE effects with a 32 mM 1 : 1 mixture of the Cbz-tripeptide derivative and the cytosine derivative in $CDCl_3$. Although numerous intramolecular NOE cross peaks were observed within the tripeptide derivative and the cytosine derivative, no intermolecular cross peaks between the two compounds were evident. Apparently intermolecular NOE effects between the peptide and the cytosine are precluded by insufficiently tight binding (K_a = 285 M^{-1}), a short lifetime for the dimolecular complex, and $T_1\rho$ values for the hydrogen bonding protons that are too small to allow observation of intermolecular cross relaxation. An effort to demonstrate an NOE effect between the peptide NH(II) proton and the cytosine $^{15}N(3)$, in a selective $^{15}N(^1H)$ experiment, was also unsuccessful.

evolutionary descendants, the nucleic acids of present organisms, and their synthesis thus presents a considerable challenge. Their specific protein binding capacities could have experimental or medical uses unrelated to their possible value in elucidating the origin of amino acid codon assignments.

ACKNOWLEDGMENTS

We are grateful for valuable help and advice to: Dr. Adriaan Bax, Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases; Dr. Jack S. Cohen, Division of Cancer Treatment, National Cancer Institute; Dr. Dennis A. Torchia, Laboratory of Biochemistry, National Institute of Dental Research; and Dr. William M. Egan, Division of Biochemistry and Biophysics, Bureau of Biologics, Food and Drug Administration.

REFERENCES

1. Woese, C. R., Dugre, D. H., Saxinger, W. C., and Dugre, S. A. (1966) Proc. Natl. Acad. Sci. U.S.A. 55, 966-974.
2. Crick, F. H. C. (1968) J. Mol. Biol. 38, 367-379.
3. Black, S. (1973) Adv. Enzymol. Relat. Areas Mol. Biol. 38, 193-234.
4. Chao, H. M., Hazel, B., and Black S. (1976) J. Biol. Chem. 251, 6924-6928.
5. Niu, C.-H., and Black S. (1981) In Peptides: Synthesis, Structure, and Function (D. H. Rich and E. Gross, eds.) pp. 327-330. Pierce Chemical Co., Rockford, Illinois.
6. Niu, C.-H., and Black, S. (1979) J. Biol. Chem. 254, 265-267.
7. Raszka, M., and Kaplan, N. O. (1970) Proc. Natl. Acad. Sci. U.S.A. 69, 2025-2029.
8. Shoup, R. R., Miles, H. T., and Becker, E. D. (1966) Biochem. Biophys. Res. Commun. 23, 194-201.
9. Schweizer, M. P., Broom, A. D., Ts'o, P. O. P., and Hollis, D. P. (1968) J. Am. Chem. Soc. 90, 1042-1055.
10. Katz, L., and Penman, S. (1966) J. Mol. Biol. 15, 220-231.
11. Iwahashi, H., and Kyogoku, Y. (1977) J. Am. Chem. Soc. 99, 7761-7765.
12. Cohen, J. S. (1973) In Experimental Methods in Biophysical Chemistry (C. Nicolau, ed.) pp. 521-588. Wiley-Interscience Publications, New York.
13. Markowski, V., Sullivan, G. R., and Roberts, J. D. (1977) J. Am. Chem. Soc. 99, 714-717.
14. Macura, S., Huang, Y., Suter, D., and Ernst, R. R. (1981) J. Mag. Res. 43, 259-281.
15. Bax, A., and Lerner, L. (1986) Science 232, 960-967.
16. Bothner-By, A. A., Stephens, R. L., Lee, J., Warren, C. D., and Jeanloz, R. W. (1984) J. Am. Chem. Soc. 106, 811-813.
17. Bax, A., and Davis, D. G. (1985) J. Mag. Res. 63, 207-213.
18. Niu, C.-H. (1984) Anal. Biochem. 139, 404-407.