ON THE CODING OF GENETIC INFORMATION

By

M. W. Nirenberg, O. W. Jones, P. Leder,

B. F. C. Clark, W. S. Sly and S. Pestka.

National Institutes of Health, Bethesda, Md.

In Press

Cold Spring Harbor Symposium 1963

The process of expressing genetic information by aligning amino acids in proper sequence during protein synthesis usually requires the RNA polymerase catalyzed synthesis of a strand of RNA complementary to DNA. Recent experiments reported at this Symposium and elsewhere suggest that mRNA (messenger RNA) synthesized in vivo is complementary to only one of the two strands of DNA (Robison and Guild, 1963; Marmur, 1963; Spiegelman, 1963; Wood and Berg, 1963). The mRNA becomes bound to ribosomes, perhaps forming a polysomal aggregate and the amino acids may be carried to these sites and ordered in correct sequence by specific transfer RNA species. It is possible that mRNA codewords are read from a fixed point by nucleotide sequences in transfer RNA complementary to those in mRNA codewords.

Thus coding errors during protein synthesis may be minimized by the requirement for correct recognition at three successive steps; that is at the DNA- mRNA, mRNA-transfer RNA (or other intermediate), and amino acid-transfer RNA- activating enzyme levels. Little is known about the mechanisms which impart specificity at the last two steps.

Some Factors Influencing the Messenger Efficiency of Synthetic Polynucleotides

The effects of base composition, catalytic ability, molecular weight, and secondary structure upon the messenger activity of synthetic polynucleotides will be considered at this time.

The messenger activity of synthetic polynucleotides may be related to its molecular weight. In E. coli extracts, poly U containing more than 100 uridylic acid residues per chain has greater template activity than smaller chains (Matthaei et al 1962), but oligo A fractions containing as few as 9-10 adenylic acid residues per chain recently have been found by Jones et al. (1963) to direct polylysine synthesis. Also, in yeast extracts oligo U

of average chain length 11 actively directs phenylalanine incorporation (Marcus et al, 1963). Although certain oligonucleotides may be degraded by nucleases more rapidly than polynucleotides and thus appear to be less efficient as templates for protein synthesis, RNA chain-length must be considered when comparing template activities of different RNA fractions.

Also, secondary structure of RNA greatley influences its messenger activity. When poly U is mixed with poly A, double- and triple- stranded helices are formed which are completely inactive in directing polyphenylalanine synthesis (Nirenberg and Matthaei, 1961). Oligo A also forms helices with poly U, and the extent of inhibition of polyphenylalanine synthesis can be correlated with oligo A chain-length and oligo A-poly U helix stability (Nirenberg et al,1963). In addition, Singer et al,(1963), have investigated a series of copolymers containing varying amounts of U and G and have found that guanine-rich polymers containing a high degree of ordered secondary structure (perhaps due to G-G interactions) also are inactive as templates for protein synthesis. These results suggest that RNA with a high proportion of helical structure may have little template activity for protein synthesis. Recent experiments have shown that poly U-poly A helices do not bind to ribosomes, and for this reason may be unable to direct protein synthesis (Cukier and Nirenberg, unpublished results). It also is possible that small, localized areas of ordered structure may serve as periods in protein synthesis.

It is difficult to compare directly the messenger efficiencies of different polynucleotide preparations because the efficiency is modified by molecular size and secondary structure. However, if the average chain length and secondary structure of different RNA preparations are assumed to be

approximately equal, the data of Table I suggest that nucleotide content may not influence greatly the overall template efficiency of mRNA. Poly U, poly UC, poly ACG and poly UACG contain 1, 8, 27, and 64 triplets respectively, and the preparations of these polynucleotides shown in Table I have been found to direct 1, 4, 9, and 10 amino acids, respectively, into protein. The essential point is that approximately the same total quantity of amino acids were directed into protein by each polynucleotide. Although these data must be interpreted with care because the same factor may not limit the incorporation rate of each amino acid, they suggest that the polynucleotide preparations may have approximately equal template efficiencies and that most nucleotide sequences may be able to code for amino acids. Although nonsense sequences may exist, thus far, none have been demonstrated definitively.

Polynucleotides containing all base combinations now have been used to direct protein synthesis in E. coli extracts. A qualitative summary of these data is presented in Table 2. Only those polynucleotides containing the minimum bases necessary to direct an amino acid into protein are shown. For example, phenylalanine is directed into protein by poly U and other U containing polymers; however, since other bases are not required, phenylalanine is listed only under poly U. Poly U, poly A, and poly C direct phenylalanine, lysine and proline, respectively, into protein. Polylysine synthesized in E. coli extracts under the direction of poly A has been found to contain 3-15 lysine residues per chain (Jones, Yaron, Sober, Heppel and Nirenberg, unpublished results). No messenger activity has been demonstrated for poly G (Matthaei, et al, 1962), but the highly ordered structure of poly G might mask template activity. However, a polynucleotide composed only of hypoxanthine

(poly I) with less secondary structure than poly C, still has not been found to direct amino acids into protein. Since hypoxanthine can replace G in RNA code words, the 2-amino group of G does not appear to be essential for coding amino acids (Basilio et al, 1962, Nirnberg and Jones, unpublished data).

Each polynucleotide composed of 2 different bases has 8 triplets, but no polymer has been found to direct more than 6 different amino acids into protein. Poly UC is unique in that it codes for only 4 amino acids, even though all UC triplets appear to function as codewords (see Coding Ratio Section). It is important to note also that polynucleotides containing only two different bases direct with great specificity almost all amino acids into protein. These findings undoubtedly reflect basic molecular characteristics of both the recognition process and the general nature of the code.

The Coding Ratio

A series of poly AC and poly UC preparations with different proportions of bases were synthesized and their activities in stimulating cell-free amino acid incorporation into protein were determined.

As shown in Fig. 1, poly AC directs the incorporation into protein of proline, histidine, threonine, asparagine, glutamine and lysine at linear rates for 15-20 minutes. Reactions were terminated after 10 minutes of incubation, while the rates of incorporation were still linear.

In Table 3 is presented an example of the data obtained for each of the five poly AC preparations tested. The theoretical proportions of the

four doublet and eight triplet permutations expected in randomly-ordered poly AC (containing by analysis, 47 percent A and 53 percent C) are shown in the first and second columns respectively. In the third and fourth columns are shown the μμmoles of each C^{14}-amino acid directed into protein by this polymer. A total of 1685 μμmoles of amino acids were directed into protein, and the relative proportions of each amino acid incorporated, in percent, are shown in the last column.

The 4 doublet permutations do not contain enough specific information to code for the 6 amino acids incorporated, whereas the information content of the 8 triplet words is adequate. The percent incorporation of lysine, asparagine, glutamine and histidine agrees well with triplet codeword frequencies, but not with doublet frequencies. If all triplets were read, some amino acids would respond to 2 or more codewords, for 6 amino acids would then be coded by 8 words. In such cases, the sum of the triplet frequencies would have to be compared with the corresponding amino acid incorporation data. For example, if CAA and CCA both coded for one amino acid, the sum of their frequencies is 24.9 percent, which cannot be distinguished from the frequency of the doublet CA (also 24.9 percent). Therefore this experimental approach may allow determination of the coding ratio for some, but not all, amino acids.

Analysis of a series of polynucleotides with varying base-ratios permits comparisons to be made with greater accuracy. The expected statistical relationship between codeword frequency and polynucleotide base-ratio are presented graphically in Figs. 2 and 3. Theoretical frequencies in percent of doublet and triplet codewords are shown on the ordinate and the base-ratio

is shown on the abscissa. Nucleotide sequence is arbitrary, and each curve represents only one of the three possible sequence permutations. As noted before, the sum of the frequencies of the triplets AAC and ACC equals the frequency of the doublet AC. Thus the AC curve represents either the doublet AC, or the sum of the two triplets AAC plus ACC. Also shown are the observed C^{14}- amino acid incorporation data. Each point represents a different poly AC preparation with the indicated base-ratio. As shown in Fig. 2, the observed incorporation of C^{14}- histidine agrees well with the theoretical frequency of the triplet ACC and differs markedly from both the AAC triplet and AC doublet curves. The data also demonstrate that the observed incorporations of both C^{14} - asparagine and C^{14}- glutamine agree well with the frequencies of AAC triplets. In contrast, the incorporation of C^{14}- threonine is similar to the expected frequencies of either the doublet AC, or the two triplets, AAC plus ACC. Therefore, threonine appears to be coded <u>either</u> by a doublet or by two triplets, and it is not possible to differentiate between these alternatives on the basis of these data.

In Fig. 3 are presented the template activities of poly AC preparations for C^{14}- proline and C^{14}- lysine. The experimentally obtained incorporation data indicate that proline is coded either by the doublet CC or by the two triplets CCC and CCA. C^{14}- lysine appears to be coded by the triplet AAA.

<u>C^{14}-Amino Acid Incorporation Directed by Poly UC.</u> The data of Fig. 4 show that proline is directed into protein either by the doublet CC or by the sum of the two triplets CCC and CCU. C^{14}-phenylalanine appears to be coded either by the doublet UU or by the two triplets UUU and UUC. It is important to note that if the codewords corresponding to these amino acids are triplets,

both CCC and CCU would code for proline and both UUU and UUC would code for phenylalanine.

In Fig. 5 are shown the poly UC-directed serine and leucine incorporation data. Both serine and leucine appear to be coded either by the doublet UC, or by the two triplets UUC and UCC. Coding of serine or of leucine by one, rather than 2 triplets is not indicated. It is important to note that if serine is coded by triplets, one triplet would have to contain 2 U residues and the other 2 C residues. Triplet words for leucine also would contain either 2 U or 2 C residues.

These experiments strongly suggest that histidine, asparagine, glutamine and lysine are coded by triplet words and that the RNA code cannot be composed only of doublets. Threonine, proline, phenylalanine, serine and leucine were found to be coded either by multiple triplets or by doublets. These data are summarized in Table 4. A mixed doublet-triplet code cannot be excluded on the basis of the available data; however, a uniform code containing only triplets would appear more probable.

The Current Codeword Dictionary. Assuming for the present that all amino acids are coded by triplets, current approximations of RNA codewords may be summarized as shown in Table 5. Nucleotide sequence is arbitrary. Fifty of the 64 possible triplets have been assigned. Almost all amino acids can be coded by polynucleotides containing 2 different bases. Since polynucleotides containing 3 bases direct protein synthesis as efficiently as polymers containing only 2 bases, it seems probable that most 3 base words are recognized. Tentative assignments are given for

such words.

It seems clear that most amino acids are coded by multiple words. Furthermore, multiple words corresponding to one amino acid often differ in base composition by only 1 nucleotide. These observations also suggest that nucleotide sequences in multiple words often may be identical. A triplet code may be constructed wherein correct hydrogen bonding between 2 out of 3 nucleotide pairs may, in some cases, suffice for coding, or alternatively, a base at one position in the triplet sometimes may pair optionally and correctly with 2 or more bases. It should be noted that a triplet code of this type in some respects would bear a superficial resemblance to a doublet code and would be in accord with all of the data available.

The coding data obtained thus far clearly indicate that most nucleotide sequences can code for amino acids with great specificity. Weisblum et al (1962) have reported that multiple species of leucine transfer RNA recognize different codewords in synthetic polynucleotides; however, additional data presented at this symposium by Benzer and by von Ehrenstein and Gonano suggest that codeword specificity in directing leucine incorporation may be greater with synthetic polynucleotides than with natural mRNA. It is important to emphasize the possibility that randomly-ordered synthetic polynucleotides may test the cell's potential to recognize codewords, and that the entire potential may not be utilized in vivo, except perhaps during mutation. Thus mRNA synthesized by a cell may not contain as many codewords as randomly-ordered polynucleotides.

Several groups of amino acids are shown in Table 6 which either are synthesized in vivo from the same precursor, or have similar structures. For example phenylalanine, tyrosine and tryptophan are derived from shikimic acid, and isoleucine, valine and leucine are synthesized from α-keto butyrate. RNA codewords corresponding to these amino acids are shown also.

Such comparisons suggest that a family of amino acids may recognize a family of codewords whose members contain similar bases. Although not all amino acids fit this pattern, enough additional examples may be cited to warrant the suggestion that such relationships reflect either the evolutionary development of the code, or the recognition of nucleotides in codewords by amino acids. The latter has been proposed by Woese (1963) and also is discussed by Weinstein in this Symposium.

OLIGODEOXYTHYMIDYLATE DIRECTED POLYLYSINE SYNTHESIS

The chemical synthesis of oligodeoxynucleotides by the method of Khorana and his associates (1961, 1962) and the demonstration of an oligodeoxynucleotide-dependent synthesis of polyribonucleotides, catalyzed by RNA polymerase (Furth et al, 1961, Stevens, 1961, Chamberlain and Berg, 1962, Falaschi et al, 1963), provided an opportunity to study their ability to stimulate cell-free amino acid incorporation. Since poly A serves as a template for polylysine synthesis (Gardner et al, 1962), oligo dT (oligodeoxythymidylate) has been used to direct poly A, and subsequent polylysine synthesis, as follows:

1) ATP $\xrightarrow[\text{RNA Polymerase}]{\text{Oligo dT}}$ Poly A + PP

2) Lysine $\xrightarrow[\underline{E.\ coli}\ \text{Extracts, etc.}]{\text{Poly A}}$ Polylysine

In addition, natural DNA and poly U have been shown to direct polylysine synthesis.

Poly A was synthesized in RNA polymerase oligo dT reaction mixtures (stage I) as described in the legend accompanying Fig. 6, and then components supporting amino acid incorporation into protein (Stage II) as in the legend of Fig. 7, were added. After further incubation, incorporation of C^{14}-lysine into polylysine was determined by precipitation with a TCA - tungstate solution (Gardner et al, 1962).

The data of Fig. 6, show that C^{14}-AMP incorporation was dependent upon the addition of oligo dT_{13-14} (13-14 nucleotides per chain) to stage I reaction mixtures, and that C^{14}-AMP incorporation was proportional to the

amount of oligo dT added within the range of 2.4 or less mmoles of nucleotide residues in oligo dT_{13-14}.

The average chain length of the C^{14}-product synthesized in the presence of oligo dT_{13-14} was determined by deproteinizing the reaction mixtures, removing the C^{14}-ATP by paper chromatography, hydrolyzing the C^{14}-product in 0.3 N KOH and separating the nucleotides by paper chromatography. The radioactivity of adenosine, adenosine-3' (2')-5'-diphosphate and adenosine-3' (2')-monophosphate was found to be 339, 308 and 22,900 counts per minute, respectively. Thus, oligo dT_{13-14} stimulated the synthesis of poly A of average chain length 60-70 pA (adenylate) residues. These data confirm similar results obtained by Furth et al (1961) and Falaschi et al (1963).

Falaschi et al (1963) also demonstrated that oligo dT chains are not elongated by the addition of (pA) residues to the free 3-hydroxyl ends of oligo dT chains, and have obtained evidence which suggests that oligodeoxynucleotides serve as templates rather than primers. Although our RNA polymerase preparations were purified 100-150 fold (Chamberlain and Berg, 1962), we have detected unprimed nucleotide incorporation under other conditions. Further enzyme purification will be necessary to determine unequivocally whether oligodeoxynucleotides function only as templates in this system.

Characteristics of the C^{14}-Polylysine Synthesis. After incubating stage I reaction mixtures at 37°, stage II components were added as described in the legend accompanying Fig. 7 . No increase in C^{14}-lysine incorporation was found in the absence of oligo dT, whereas the addition of 1.2 mmoles of (pdT) residues in oligo dT_{13-14} stimulated C^{14}-lysine incorporation at a

C^{14}-AMP incorporation into poly A was determined at the end of the stage I incubation. No incorporation was observed in the absence of oligo dT, whereas the addition of 1.2 nmoles of pdT in oligo dT directed the incorporation of 3.8 nmoles of AMP into poly A. Poly A, in turn, directed the incorporation of 0.64 nmoles of C^{14}-lysine into polypeptide.

The effect of oligo dT upon the individual incorporation of 18 other C^{14}- amino acids (minus cysteine) also was determined. Only C^{14}-lysine was directed into protein, which demonstrated marked amino acid specificity.

Oligo dT was found to be resistant to digestion with pancreatic RNase. RNase partially inhibited C^{14}-lysine incorporation into protein, whereas, pazoryein and pancreatic RNase were strongly inhibitory. As expected, no inhibition was observed in the presence actinomycin D.

It is important to note that if stage I incubation was not performed, oligo dT directed little or no C^{14}-AMP incorporation into poly A in stage II reaction mixtures, and thus no stimulation of C^{14}-lysine incorporation could be observed.

RNA polymerase has been found to catalyze the synthesis of complementary RNA in the presence of either RNA or DNA templates (Weiss and Nakamoto, 1961, Weiss, 1963, Krakow and Ochoa, 1963).

Poly A synthesized from poly U templates was found to stimulate C^{14}-lysine incorporation into protein. In other experiments no stimulation of C^{14}-lysine incorporation was observed when poly C rather than poly U

was added to stage I reaction mixtures. RNA polymerase also has been shown (Chamberlain and Berg, 1962) to catalyze a DNA dependent synthesis of poly A from ATP (in the absence of UTP, GTP and CTP). Under these conditions, poly A synthesized under the direction of calf thymus DNA stimulated C^{14}-lysine incorporation.

Paper chromatography of polylysine as described in the legend accompanying Fig. 8, permits separation of lysine peptides of different chain lengths. Peptides containing approximately eleven or more lysine residues remain at the origin, whereas the mobilities of smaller peptides are as follows: lysine > di- > tri- > tetra- > penta- > hexa- > hepta- > octa- > nona- > deca-lysine. As shown in Fig. 8, most of the C^{14}-product synthesized in the presence of oligo dT_{13-14} remained at the origin after chromatography. Digestion with trypsin converted the C^{14}-product almost quantitatively to peptides which migrated with free, di-, tri-, and tetra-lysine characteristics.

In separate experiments the C^{14}-product was eluted from the origin. An aliquot was hydrolyzed completely with HCl, and another, with trypsin. Both aliquots were chromatographed as before. One C^{14}-spot having the characteristic mobility of free lysine was found following acid hydrolysis. After digestion with trypsin, C^{14}-products with the expected mobilities of free, di-, tri-, and tetra-lysine were found. In addition the C^{14}-polylysine which remained at the origin was shown to contain carboxyl-terminal C^{14}-lysine residues by a hydrazinolysis method (Akabori et al, 1952).

Effect of Molecular Weight upon the Activity of Oligo dT. Falaschi et al. (1963) demonstrated that oligo dT chains containing less than 4 residues

AKABORI, S., K. OHNO, and K. NARITA. 1952. On the hydrazinolysis of proteins and peptides. Bull. Chem. Soc. Japan, 25: 214-218.

BASILIO, C., A. J. WAHBA, P. LENGYEL, J. F. SPEYER and S. OCHOA. 1962. Synthetic polynucleotides and the amino acid code, V. Proc. Nat. Acad. Sci., U.S., 48: 613-616.

CHAMBERLAIN, M. and P. BERG. 1962. DNA directed synthesis of RNA by an enzyme from E. coli. Proc. Nat. Acad. Sci. U.S., 48: 81-94.

FALASCHI, A., J. ADLER and H. G. KHORANA. 1963. Chemically synthesized deoxypolynucleotides as templates for RNA polymerase. J. Biol. Chem. (in press).

FURTH, J. J., J. HURWITZ and M. GOLDMANN. 1961. The directing role of DNA in RNA synthesis. Biochem. Biophys. Res. Com., 4: 362-367.

GARDNER, R.S., A.J. WAHBA, C. BASILIO, R. S. MILLER, P. LENGYEL and J. F. SPEYER. 1962. Synthetic polynucleotides and the amino acid code, VII. Proc. Nat. Acad. Sci. U.S., 48: 2087-2093.

JONES, O. W., E. E. TOWNSEND, H. A. SOBER and L. A. HEPPEL. 1963. Effect of chain length on the template activity of polyribonucleotides. Biochem. (in press).

KHORANA, H. G. and J. P VIZSOLYI. 1961. Studies on polynucleotides VIII. J. Am. Chem. Soc., 83: 675-685.

KHORANA, H. G. and J. P. VIZSOLYI. 1962. Studies on polynucleotides XII. J. Am. Chem. Soc., 84: 414-419.

KRAKOW, J. S. and S. OCHOA. 1963. RNA polymerase of Azotobacter vinelandii, I. Proc. Nat. Acad. Sci. U.S., 49: 88-94.

MARCUS, L., B. K. BRETTHAUER, R. N. BOCK and H. O. HALVORSON. 1963. The effect of poly U size on the incorporation of phenylalanine in the cell-free yeast system. Proc. Nat. Acad. Sci. U.S., (in press).

MARMUR, J. 1963. Biological and physical properties of bacillus bacteriophage DNA. This Symposium.

MATTHAEI, J. H., O. W. JONES , R. G. MARTIN, and M. W. NIRENBERG. 1962. Characteristics and composition of RNA coding units. Proc. Nat. Acad. Sci. U.S., 48: 666-677.

NIRENBERG, M. W. and J. H. MATTHAEI. 1961. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proc. Nat. Acad. Sci. U. S., 47: 1588-1602.

NIRENBERG, M. W., J. H. MATTHAEI, O. W. JONES, R. G. MARTIN and S. H. BARONDES. 1963. Approximation of the genetic code via cell-free protein synthesis directed by template RNA. Fed. Proc. 22: 55-61.

ROBISON, M. and W. B. GUILD. 1963. Evidence for messags reading from a ubique strand of DNA. Fed. Proc., 22: 643.

SINGER, M.F., O.W. JONES and M. W. NIRENBERG. 1963. The effect of secondary structure on the template activity of polyribonucleotides. Proc. Nat. Acad. Sci. U.S., 49 392-399.

SPIEGELMAN, S. 1963. Some properties of the genetic transcription mechanism. This Symposium.

STEVENS, A. 1961 Net formation of polyribonucleotides with base compositions analogous to DNA. J. Biol. Chem., 236: PC43-PC45.

WALEY, S.G. and J. WATSON. 1953. The action of trypsin on polylysine. Biochem. J.,55: 328-337.

NAIRN, A. J., R. S. GARDNER, C. BASILIO, H. S. MILLER, J. F. SPRYER and P. LENGYEL. 1963. Synthetic polynucleotides and the amino acid code, VIII. Proc. Nat. Acad. Sci. U.S., 49: 116-122.

WEISS, S. B. and T. NAKAMOTO. 1961. On the participation of DNA in RNA synthesis. Proc. Nat. Acad. Sci. U.S., 47: 694-697.

WEISS, S. B. 1963. Properties of DNA-dependent synthesized RNA. pp. 51-67. Informational Macromolecules. New York. Academic Press.

WEISBLUM, B., S. BENZER and R. W. HOLLEY. 1962. A physical basis for degeneracy in the amino acid code. Proc. Nat. Acad. Sci. U.S., 48: 1449-1454.

WOESE, C. 1963. The Genetic Code. I.C.S.U. Reviews, (in press)

WOOD, W.B. and P. BERG. 1963. Studies on the "messenger" activity of RNA synthesized with RNA polymerase. This Symposium.

TABLE 1. TEMPLATE ACTIVITIES OF 1, 2, 3 AND 4 BASE POLYNUCLEOTIDES

POLYNUCLEOTIDE		U	UC	ACG	UACG	
	U	100	47	-	56	
BASE RATIO	C	-	53	32	13	
MOLES-PERCENT	A	-	-	46	5	
	G	-	-	22	25	
POSSIBLE TRIPLETS		1	8	27	64	
C^{14}-AMINO ACIDS DIRECTED INTO PROTEIN		PHE	PHE	LYS	ILEU	PRO
			LEU	ALA	MET	LYS
			SER	ARG	CYSH	ALA
			PRO	SER	VAL	ARG
				THR	GLY	THR
				GLU-NH_2	TRY	GLU-NH_2
				ASP-NH_2	TYR	ASP-NH_2
				HIS	PHE	HIS
				PRO	LEU	
					SER	
TOTAL C^{14}-AMINO ACID INCORPORATION (mμmoles)		3.21	2.91	2.22	5.09	

Legend for Table 1.

See text for details.

Codeword nucleotide sequences are arbitrary.

TABLE 2. SUMMARY OF CODING DATA

POLY	U	A	C	G
	PHE	LYS	PRO	-

POLY	UA	UC	UG	AC	AG	CG	UAG
	TYR	LEU	LEU	HIS	ARG	ARG	MET
	LEU	SER	VAL	ASP-NH_2	GLU	ALA	ASP
	ILEU		CYSH	GLU-NH_2	GLU-NH_2	SER	
	ASP-NH_2		TRY	THR	ASP*	THR*	

Legend for Table 2.

Only those polynucleotides containing the minimal number of bases necessary to stimulate an amino acid into protein are shown. Amino acids coded by homopolynucleotides are not listed again under randomly-ordered polynucleotides.

*Predicted

' Reported by Wahba et al (1963)

TABLE 3

COMPARISON BETWEEN AMINO ACIDS INCORPORATED AND RNA CODEWORD FREQUENCIES

Theoretical Codeword Frequency in Poly AC Containing 47% A and 53% C		C^{14}-Amino Acids Incorporated into Protein		
DOUBLETS	TRIPLETS	C^{14}-AMINO ACID	μμMOLES INCORPORATED	FREQUENCY OF C^{14}-AMINO ACIDS INCORPORATED
percent				percent
AA 22.1	AAA 10.4	LYSINE	183	10.8
AC 24.9	AAC 11.7	ASPARAGINE	192	11.6
CA 24.9	ACA 11.7	GLUTAMINE	157	9.3
CC 28.1	CAA 11.7	THREONINE	444	26.3
Total 100.0	CCA 13.2	HISTIDINE	159	9.4
	ACC 13.2	PROLINE	550	32.6
	CAC 13.2			
	CCC 14.9		Total 1685	Total 100.0
	Total 100.0			

LEGENDS FOR TABLES

Legend for Table 3 [illegible] Each reaction mixture contained the following components in a final volume of 0.25 ml: 0.1 M Tris, pH 7.8; 0.01 M magnesium acetate; 0.05 M KC l; 6×10^{-3} M mercaptoethanol; 1×10^{-3} M ATP; 5×10^{-3} M potassium phosphoenolpyruvate; 5 μg of crystalline phosphoenolpyruvate kinase (Calif. Corp. Biochem. Research); 0.8×10^{-4} M C^{14}-amino acid; (approximately 30,000-150,000 counts/minute/reaction mixture); 3.2×10^{-4} M each of 19 C^{12}-L amino acids minus the C^{14}-amino acid; 15 μg of polynucleotide when specified; and 1.1 mg *E. coli* preincubated S-30 protein (Nirenberg and Matthaei, 1961) Reaction mixtures were incubated at 37° for 10 minutes. Protein precipitation, washing, and counting were performed as described by Nirenberg and Matthaei (1961).

The theoretical frequencies in percent, of doublets and triplets in polynucleotides were calculated as follows: The frequency of the triplet AAA in this poly AC preparation would be .47 x .47 x .47 x 100 = 10.4 percent. The doublet frequency for CA would be .47 x .53 x 100 = 24.9 percent.

The μmoles of each amino acid incorporated in the absence of polynucleotide were: lysine, 30; asparagine, 54; glutamine, 49; threonine, 40; histidine, 26; proline, 36.

TABLE 4

SUMMARY OF CODING RATIO DATA

C^{14}-AMINO ACID	CODEWORD* TRIPLET	DOUBLET
HISTIDINE	ACC	-
ASPARAGINE	CAA	-
GLUTAMINE	AAC	-
LYSINE	AAA	-
THREONINE	CCA + ACA	or AC
PROLINE	CCC + CAC + CUC	or CC
PHENYLALANINE	UUU + UCU	or UU
SERINE	CUU + CCU	or CU
LEUCINE	UUC + UCC	or UC

* Nucleotide sequences are arbitrary.

TABLE

SUMMARY OF RNA CODEWORDS

AMINO ACID	RNA CODE WORDS *			
ALANINE	CCG	UCG'	ACG'	
ARGININE	CGC	AGA	UGC'	CGA'
ASPARAGINE	ACA	AUA	ACU'	
ASPARTIC ACID	GUA	GCA'	GAA'	
CYSTEINE	UCG			
GLUTAMIC ACID	GAA	GAU'	GAC'	
GLUTAMINE	AAC	AGA	AGU'	
GLYCINE	UGG	AGG	CGG	
HISTIDINE	ACC	ACU'		
ISOLEUCINE	UAU	UAA		
LEUCINE	UUG	UUC	UCC	UUA
LYSINE	AAA	AAU		
METHIONINE	UGA			
PHENYLALANINE	UUU	CUU		
PROLINE	CCC	CCU	CCA	CCG'
SERINE	UCU	UCC	UCG'	AGC
THREONINE	CAC	CAA		
TRYPTOPHAN	GGU			
TYROSINE	AUU			
VALINE	UGU	UGA'		

* Arbitrary nucleotide sequence.
' Probable

TABLE 6. RELATIONSHIP BETWEEN AMINO ACIDS OF SIMILAR METABOLIC ORIGIN OR STRUCTURE AND THEIR RNA CODEWORDS

AROMATIC AMINO ACIDS		DICARBOXYLIC AMINO ACIDS AND AMIDES		ILEU, VAL, LEU FAMILY	
PHE	UUU	$ASP-NH_2$	AAU	ILEU	UUA
	UUC		AAC		UAA
TYR	UUA	ASP	AUG	VAL	UUG
TRY	UGG	GLU	AUG	LEU	UUG
			AAG		UUA
					UUC
		$GLU-NH_2$	AUG		UCU
			AAC		

TABLE 7

RELATION BETWEEN OLIGO dT CHAIN LENGTH AND ACTIVITY

ADDITION	C^{14}-AMP INCORPORATION (mμmoles)	C^{14}-LYSINE INCORPORATION (mμmoles)
NONE	0.1	0.075
17 mμmoles oligo dT_{6-7}	1.5	0.237
17 mμmoles oligo dT_{7-8}	6.8	0.387
17 mμmoles oligo dT_{9-11}	13.0	0.406

The components of the reaction mixtures and conditions are described in the legend accompanying Fig. 5 and 7

The rate of C^{14} amino acid incorporation into protein directed by poly AC (base ratio = A, 47 percent and C, 53 percent). Reaction mixture components are described in the legend of Table 3.

Incubations were stopped at the times indicated by the addition of 3.0 ml. of 10 percent TCA at 3°. The samples were heated at 90° for 20 minutes, chilled and then filtered through Millipore filters and washed with 5 percent TCA at 3°. Radioactivity measurements were performed in a thin-window gas flow Nuclear Chicago Corp. counter with a counting efficiency of 23 percent. Each point represents the pmoles of C^{14}-amino acid incorporated into protein due to the addition of poly AC.

Fig. 3

Comparison of C^{14}-histidine, C^{14}-threonine, C^{14}-asparagine and C^{14}-glutamine incorporation data with the theoretical frequencies of doublet and triplet codewords in poly AC preparations. The solid lines represent the experimentally determined incorporation data calculated as described in Table I. All assays were performed as described in Table 3.

Fig. 3

Comparison of C^{14}-proline and C^{14}-lysine incorporation data with the theoretical frequencies of doublet and triplet codewords in poly AC preparations. The solid lines represent the experimentally determined incorporation data calculated as described in Table 1. All assays were performed as described in Table 3.

Fig. 4

Comparison of C^{14}-proline and C^{14}-phenylalanine incorporation data with the theoretical frequencies of doublet and triplet codewords in poly UC preparations. The solid lines represent the experimentally determined incorporation data calculated as described in Table 1. All assays were performed as described in Table 3.

Fig. 5

Comparison of C^{14}-serine and C^{14}-leucine incorporation data with the theoretical frequencies of doublet and triplet codewords in poly UC preparations. The solid lines represent the experimentally determined incorporation data calculated as described in Table 1. All assays were performed as described in Table 3.

Fig. 6

Characteristics of C^{14}-poly A synthesis in RNA polymerase (stageI) reaction mixtures. In the figure on the left, reaction mixtures were incubated at 37° for 15 minutes, then were deproteinized and washed with 5 percent TCA at 3°. The symbols in the figure on the right represent the following: Δ, - polymer; ●, + 1.2 mμmoles of base residues in oligo dT_{13-14}.

Each stage I reaction mixture contained the following in a final volume of 0.125 ml: 4×10^{-2} M Tris, pH 7.6; 4×10^{-3} M $MgCl_2$; 10^{-3} M $MnCl_2$; 1.2×10^{-2} M mercaptoethanol; 1.6×10^{-3} M 8-C^{14}-ATP, tetralithium salt (Schwarz BioResearch, Inc.); and 20 μg *E. coli* RNA polymerase protein (20 units (Chamberlain and Berg, 1962)).

Fig. 7

Characteristics of oligo dT_{13-14} directed synthesis of C^{14}-polylysine. The symbols represent the following: ■ , minus oligo dT; ▲, plus 1.2 mμmoles (pdT) residues in oligo dT_{13-14}; ○ plus 2.4 mμmoles (pdT) residues in oligo dT_{13-14}; ●, plus 4.8 mμmoles (pdT) residues in oligo dT_{13-14}. Components of stage I reaction mixtures are as noted in the legend of Fig. 6. Stage II reaction mixtures contained, in 0.25 ml: 6×10^{-2} M Tris, pH 7.8, 2×10^{-3} M $MgCl_2$; 1.2×10^{-2} M magnesium acetate; 5×10^{-4} M $MnCl_2$; 1.2×10^{-2} M mercaptoethanol; 2.8×10^{-3} M ATP; 5×10^{-2} M KCl; 5×10^{-3} M potassium phosphoenolpyruvate; 5 μg crystalline phosphoenolpyruvate kinase (Calif. Corp. Biochem. Research); 2×10^{-4} M each of 19 L-amino acids; 2×10^{-4} M C^{14}-L-lysine (Nuclear Chicago Corp.) with specific radioactivity of 4-8 mμcuries/mμmole; 20 μg RNA polymerase protein (20 units (Chamberlain and Berg, 1962)) and 1.1 mg *E. coli* extract protein (Nirenberg and Matthaei, 1961). Stage I reaction mixtures were incubated at 37° for 15 minutes before the addition of stage II components.

Fig. 8

Chromatographic analysis of the C^{14}-polylysine synthesized in stage I plus stage II reaction mixtures under the direction of oligo dT_{13-14} before and after tryptic digestion. The symbols represent the following: O, minus oligo dT; ●, plus oligo dT_{13-14}; Δ, plus oligo dT_{13-14} followed by tryptic digestion. Reactions were carried out as described in the legends of Figures 6 and 7. After incubating stage II reaction mixtures at 37° for 30 minutes, the reactions were terminated by the addition of 2 ml of cold 10% TCA. The supernatant solution was extracted three times with equal volumes of ether, concentrated in vacuo, and the residue was compared chromatographically with a partial tryptic digest of chemically synthesized C^{12}-polylysine (YEDA Research and Development Co., Ltd. Rehovoth, Israel) on Whatman 3MM paper in a solvent similar to that described by Waley and Watson (1953) containing pyridine/n-butyl alchol/acetic acid/water = 6/9/3/7 v/v for 36 hours.

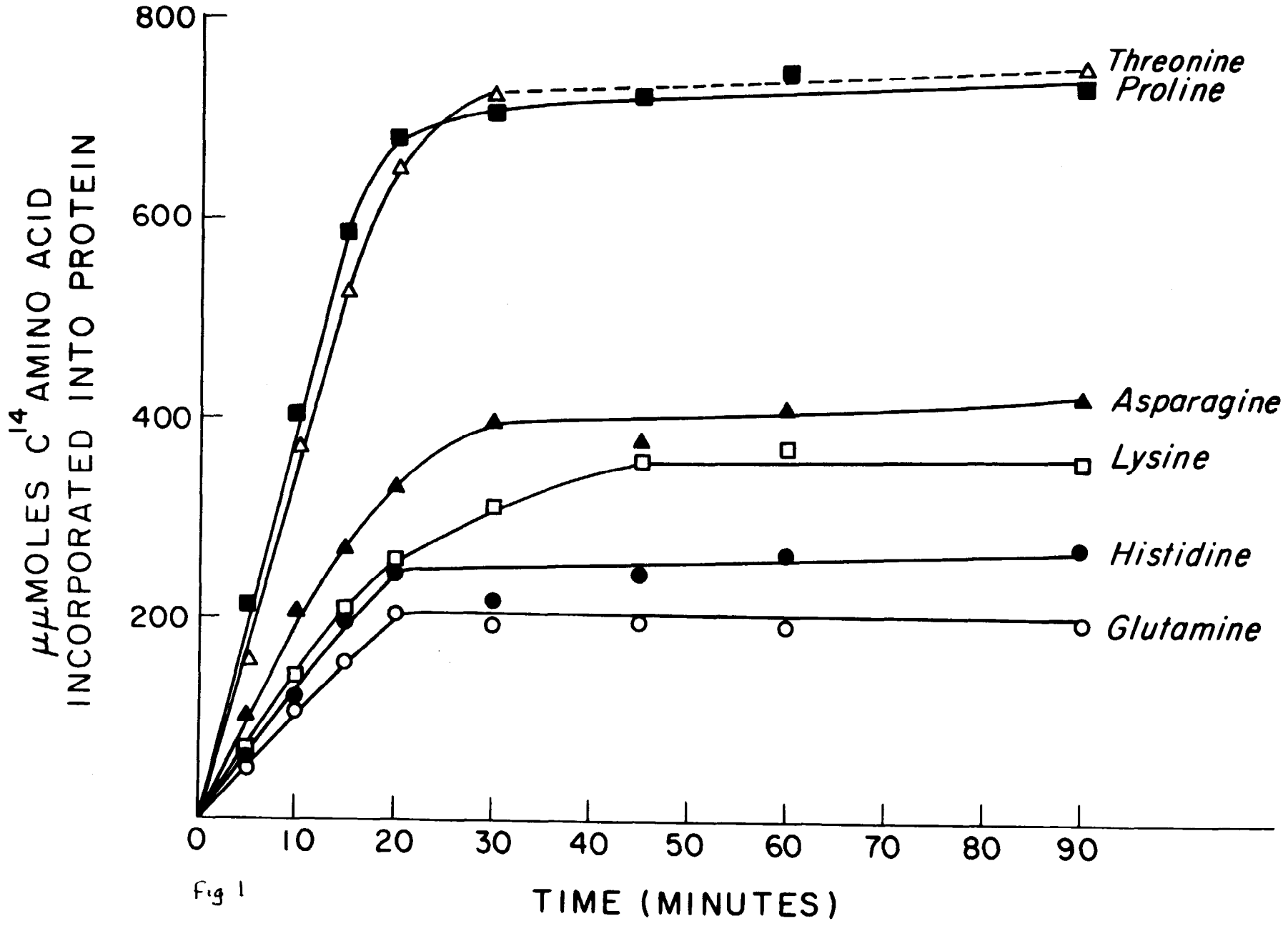
800
600
400
200
0
10
20
30
40
50
60
70
80
90
μμMOLES C14 AMINO ACID INCORPORATED INTO PROTEIN
TIME (MINUTES)
Threonine
Proline
Asparagine
Lysine
Histidine
Glutamine

Fig 1

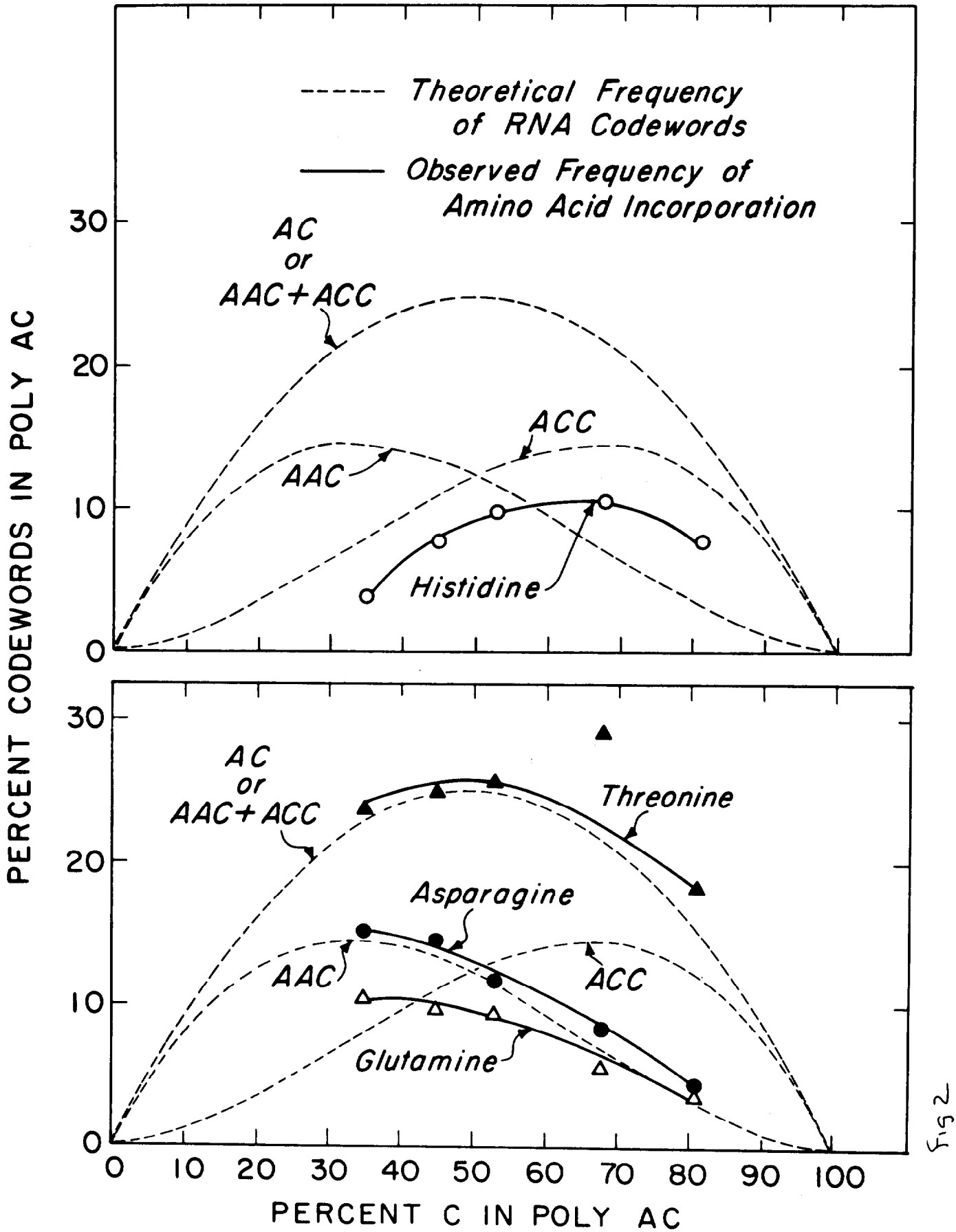
Theoretical Frequency of RNA Codewords
Observed Frequency of Amino Acid Incorporation
AC or AAC+ACC
ACC
AAC
Histidine
Threonine
Asparagine
Glutamine
PERCENT CODEWORDS IN POLY AC
PERCENT C IN POLY AC
0
10
20
30
40
50
60
70
80
90
100

Fig 2

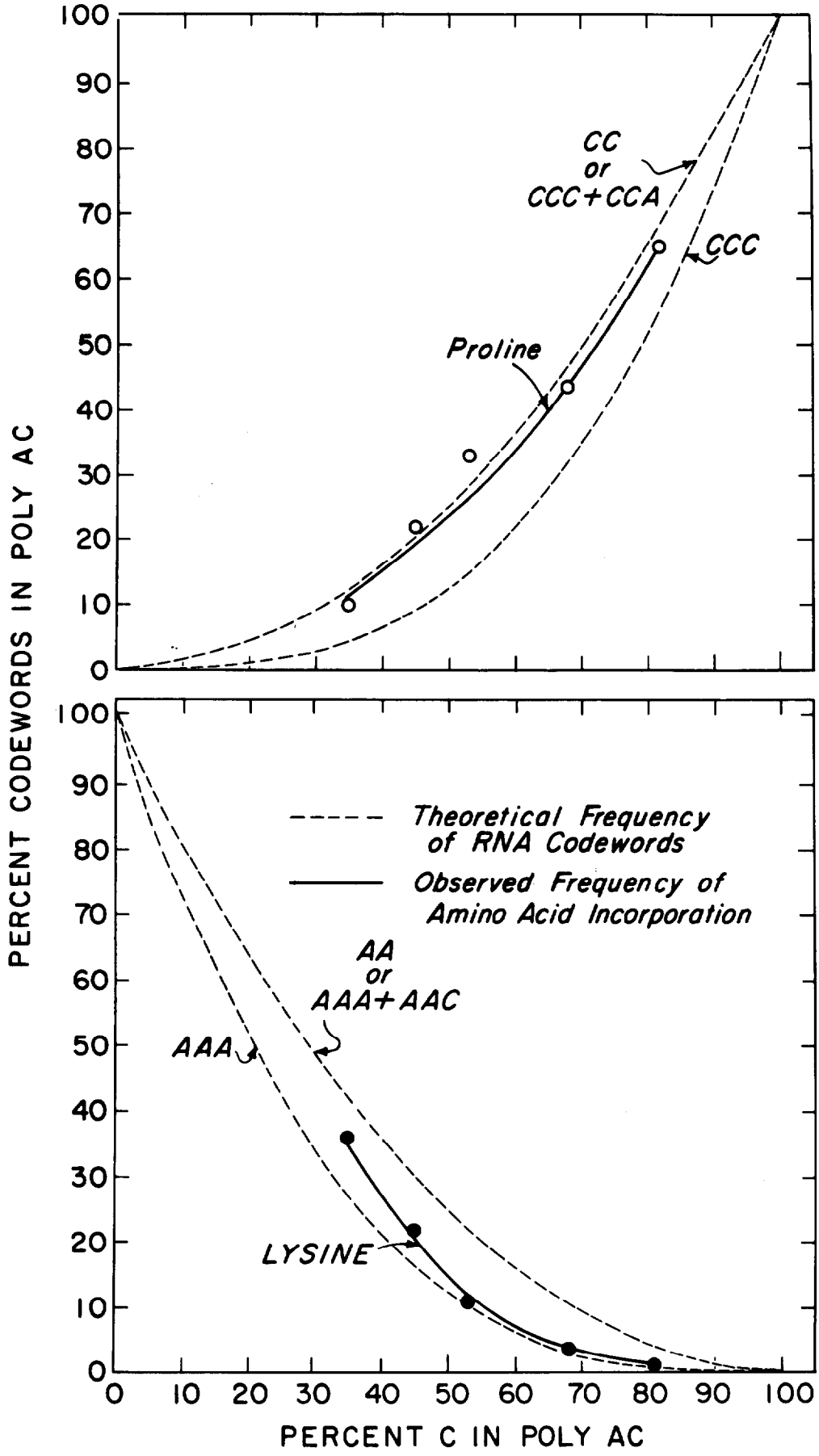
PERCENT CODEWORDS IN POLY AC
PERCENT C IN POLY AC
CC
or
CCC+CCA
CCC
Proline
Theoretical Frequency
of RNA Codewords
Observed Frequency of
Amino Acid Incorporation
AA
or
AAA+AAC
AAA
LYSINE
0
10
20
30
40
50
60
70
80
90
100

Fig. 3

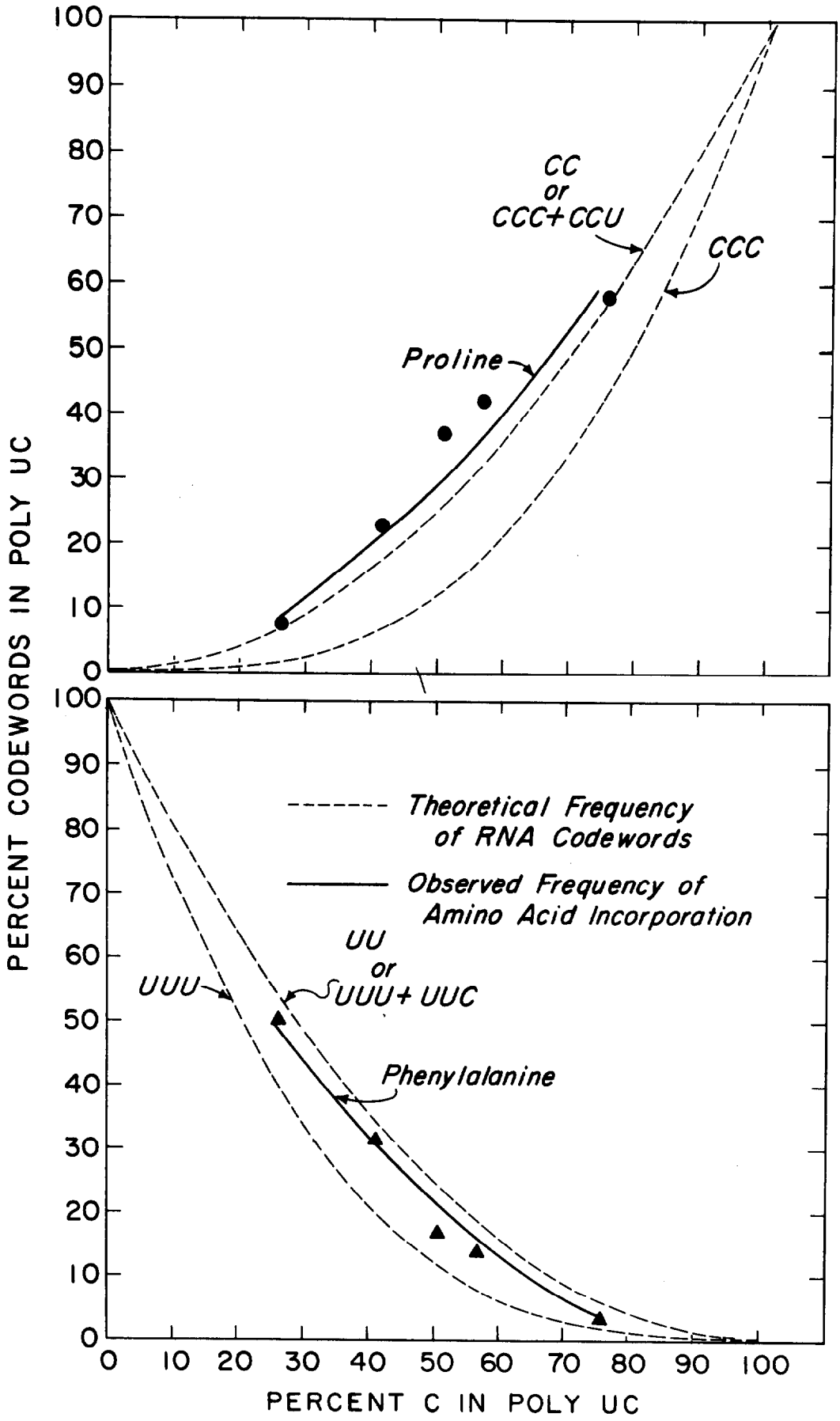
PERCENT CODEWORDS IN POLY UC
PERCENT C IN POLY UC
CC
or
CCC+CCU
CCC
Proline
Theoretical Frequency
of RNA Codewords
Observed Frequency of
Amino Acid Incorporation
UU
or
UUU+UUC
UUU
Phenylalanine

Fig 4

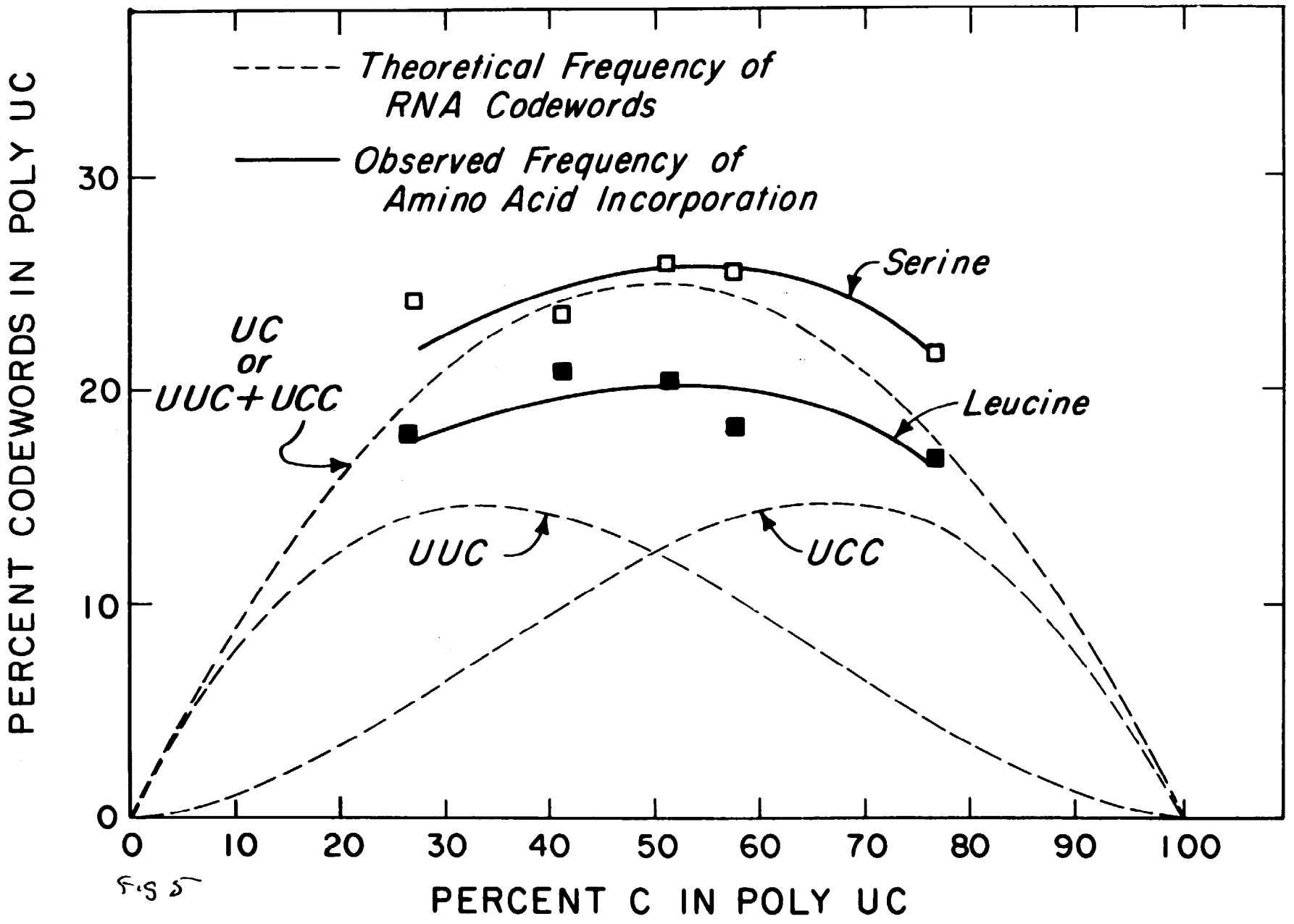
Theoretical Frequency of RNA Codewords
Observed Frequency of Amino Acid Incorporation
Serine
Leucine
UC or UUC+UCC
UUC
UCC
PERCENT CODEWORDS IN POLY UC
PERCENT C IN POLY UC
0
10
20
30
40
50
60
70
80
90
100

Fig 5

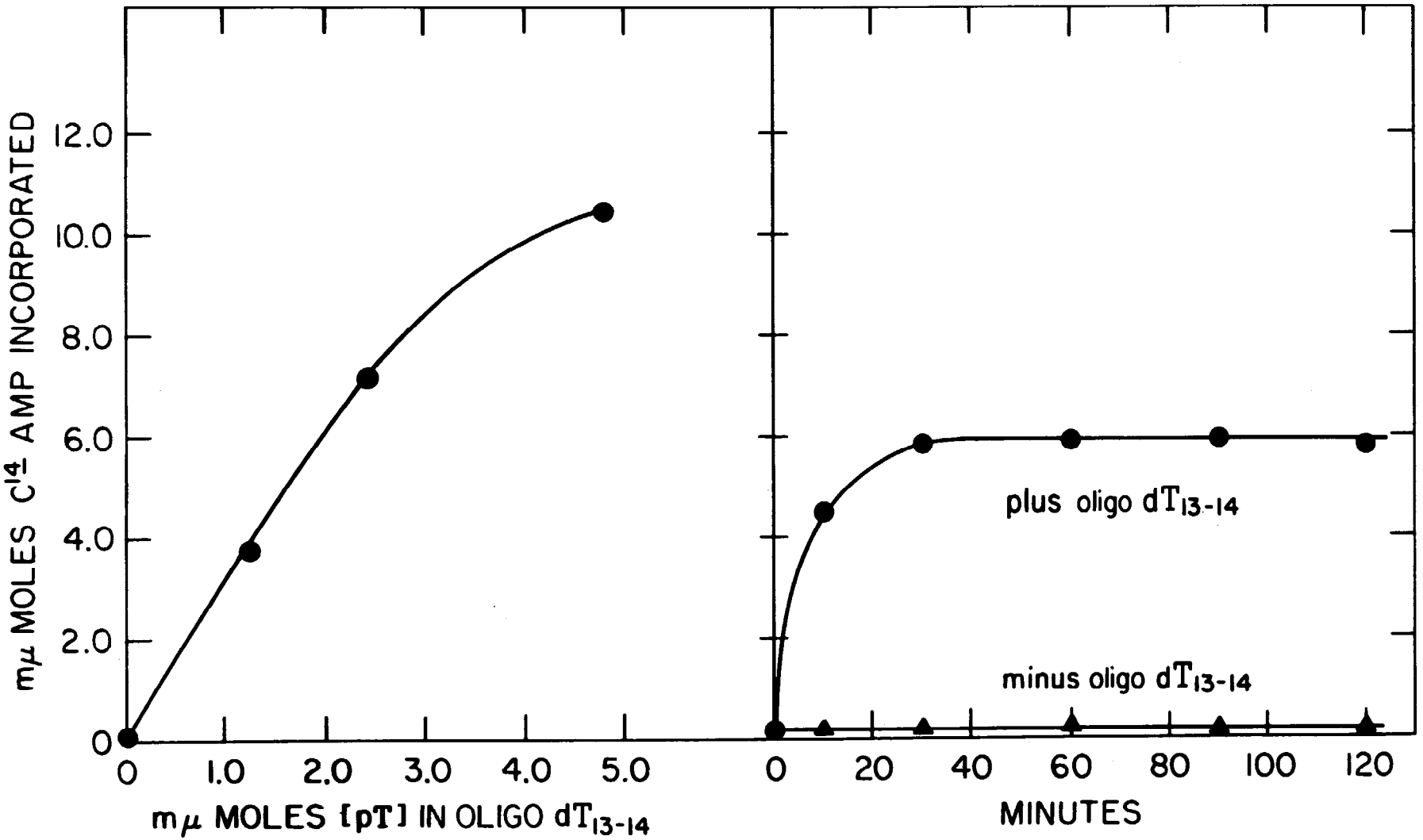
mμ MOLES C14 AMP INCORPORATED
12.0
10.0
8.0
6.0
4.0
2.0
0
0
1.0
2.0
3.0
4.0
5.0
mμ MOLES [pT] IN OLIGO dT13-14
plus oligo dT13-14
minus oligo dT13-14
0
20
40
60
80
100
120
MINUTES

Fig 6

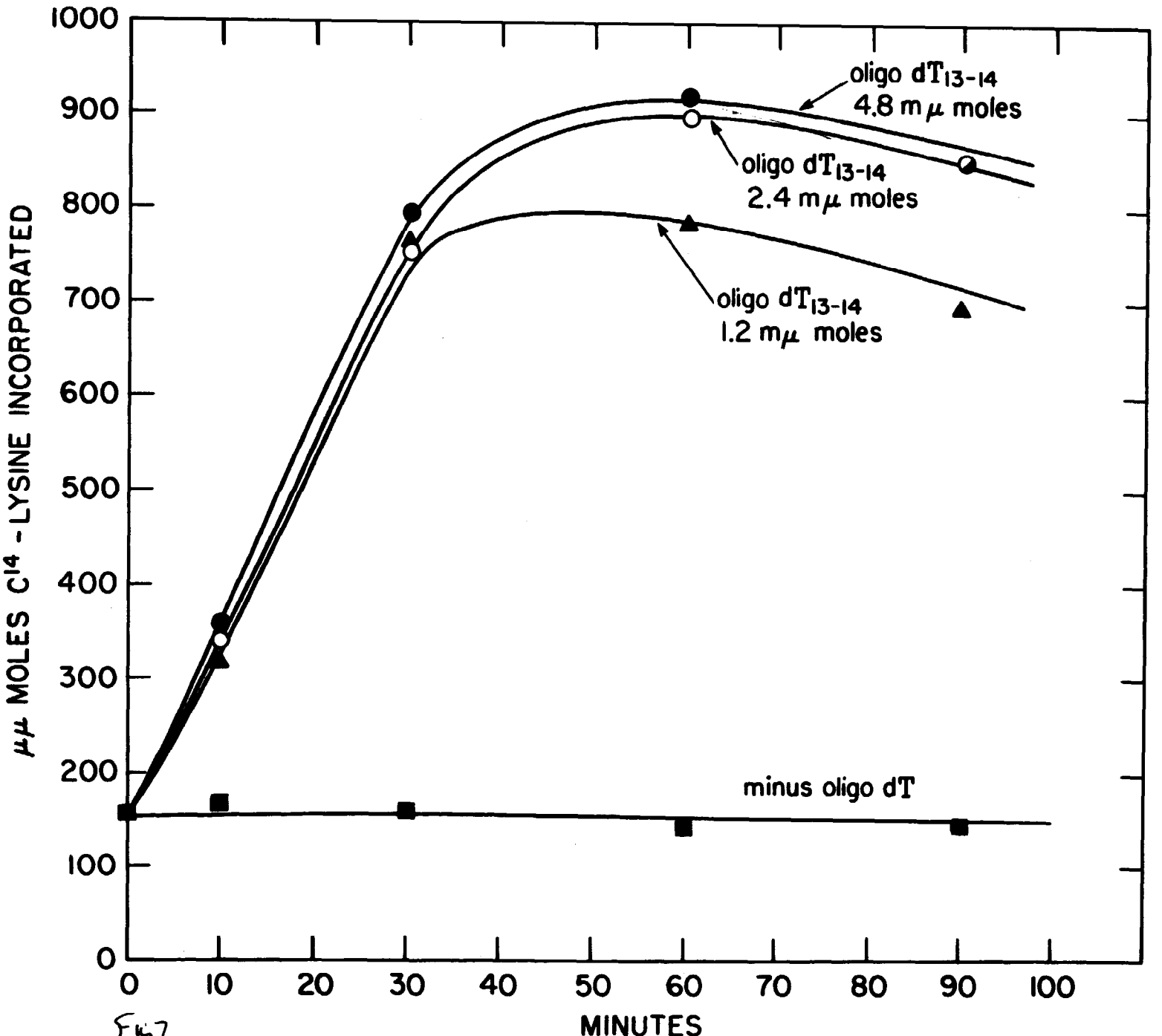
1000
900
800
700
600
500
400
300
200
100
0
μμ MOLES C14 - LYSINE INCORPORATED
0
10
20
30
40
50
60
70
80
90
100
MINUTES
oligo dT13-14
4.8 mμ moles
oligo dT13-14
2.4 mμ moles
oligo dT13-14
1.2 mμ moles
minus oligo dT

Fig 7

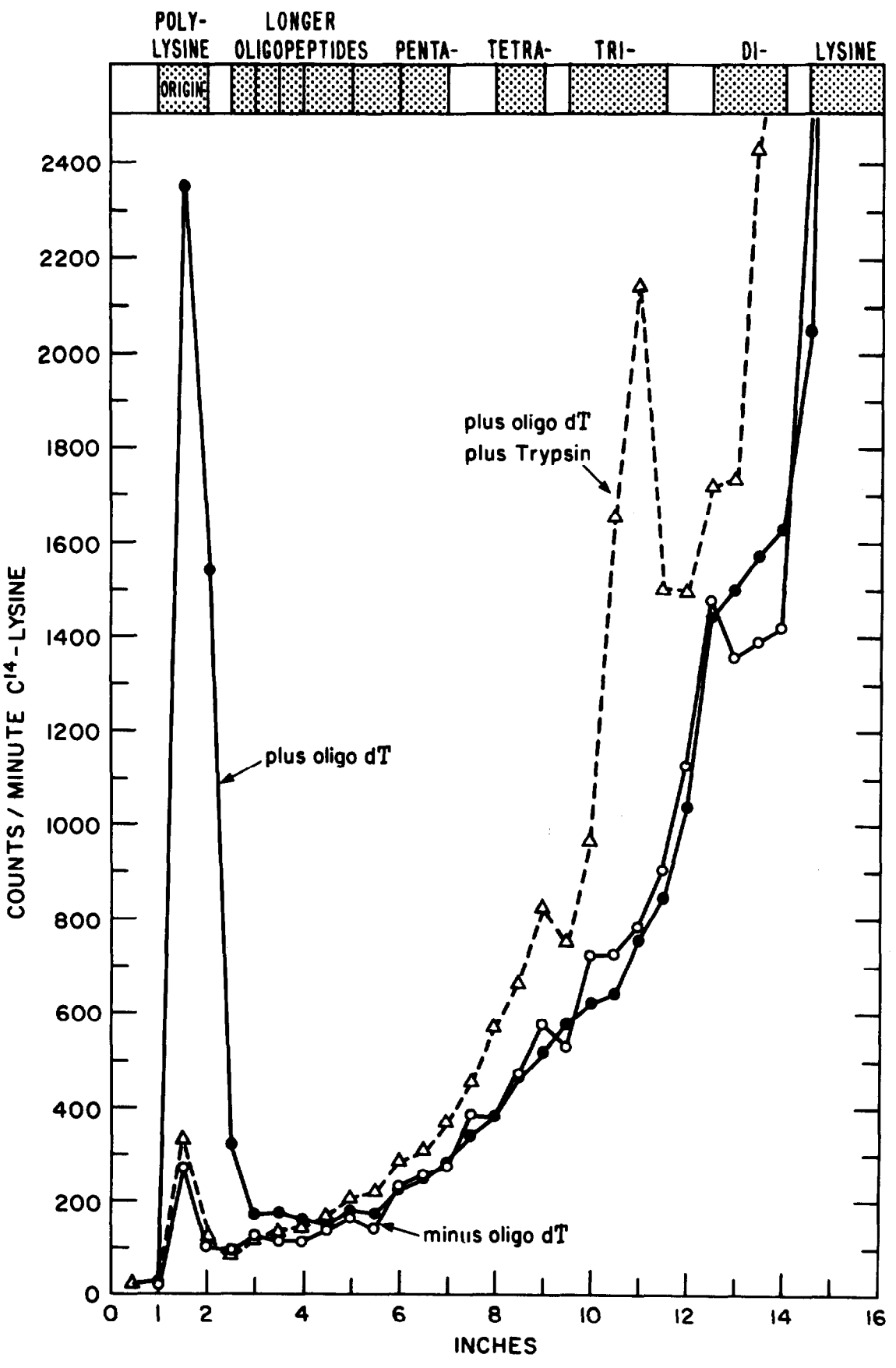
CHROMATOGRAPHIC POSITION OF LYSINE PEPTIDES
POLY-LYSINE
ORIGIN
LONGER OLIGOPEPTIDES
PENTA-
TETRA-
TRI-
DI-
LYSINE
COUNTS / MINUTE C14-LYSINE
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
plus oligo dT
plus Trypsin
plus oligo dT
minus oligo dT
0
1
2
4
6
8
10
12
14
16
INCHES

Fig 8