ARE THERE STRUCTURAL ANALOGIES BETWEEN AMINO ACIDS AND NUCLEIC ACIDS?

LAWRENCE B. HENDRY and EDWIN D. BRANSOME, Jr.

Departments of Medicine and Endocrinology, Medical College of Georgia, Augusta, Georgia 30912, U.S.A.

and

MATHEW PETERSHEIM

Department of Chemistry, University of Rochester, Rochester New York, 14623, U.S.A.

(Received 22 August, 1980)

Abstract. Space-filling molecular models have been used to examine structural analogies between amino acids and nucleic acids. The three-dimensional structures of amino acid *R* groups appear to be stereochemically related to cavities formed by removal of single bases in double helical nucleic acids. The common L amino acids may thus be complementary to their codons.

1. Introduction

Knowledge of modern chemistry provides relatively few constraints on our ability to conceive of and to chemically synthesize an almost infinite number of molecules with diverse structural frameworks. In contrast, the variety of molecules synthesized *in vivo* by the myriads of different organisms is severely limited. Moreover, the same discreet set of biochemicals can be found in many different phyla. While similarities in genetic origin must be responsible for similar constraints on biosynthesis, the question of why this is so has not been answered. The idea that nucleic acids (DNA and RNA) are the exclusive biochemicals in all species which serve as the genetic templates for the synthesis of enzyme proteins which in turn stereochemically limit the synthesis of other biologically active molecules is generally accepted. The concept that the sequence of bases in nucleic acids dictates the amino acid sequence of proteins through a phylogenetically common triplet code is also no longer questioned.

A missing piece in this biochemical puzzle is the nature of the constraints imposed by the nucleic acids which result in the incorporation of only twenty structural units, i.e., the common L-amino acids, during protein synthesis. We presume these constraints to be stereochemical. Many investigators have considered, and some have sought to establish, physicochemical and stereochemical bases for the genetic code since Gamov's (1954) first attempts*. Some of the hypotheses have been criticized for errors in the construction of molecular models and for lack of chemical detail. Their apparent lack of success can be epitomized by the idea advanced by Crick (1968) that the genetic code could be a 'frozen accident' and hence unrelated to any stereoche-

* Their publications are indicated by an asterisk in the references at the end of this paper.

Origins of Life **11** (1981) 203–221. 0302–1688/81/0113–0203$02.85.

mical relationship between nucleic acids and amino acids. We have been encouraged to take up this implicit challenge by the pioneering efforts of Woese (1961, 1965a, b, 1967, 1968, 1970) and by several recent findings of Weber and Lacey (1978) and Jungck (1978) who have reported experimental data that support a physiochemical hypothesis for the origin of the genetic code.

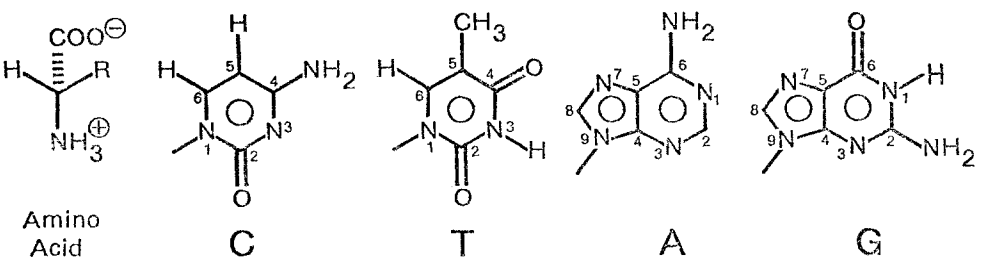
In several publications (Hendry *et al.*, 1977; Witham *et al.*, 1978; Hendry and Witham, 1979) we have reported on evidence that some biologically active small molecules – products of enzymatic biosynthesis such as steroids and phytohormones – reflect the structure of DNA or RNA in size, shape, and absolute chirality. In this paper, the first of several we intend to publish on the subject, we have directly compared the structures of amino acids to the structure of DNA, taking an approach which has to our knowledge not been used by any of our predecessors. Namely, amino acids were substituted for bases in right handed double helical DNA. We intend to show that there are structural similarities of amino acid *R* groups to purine and pyrimidine bases and that by means of stereochemical 'fits' and measurements with Corey–Pauling–Koltun (CPK) models we can make assignments of amino acids to DNA bases which are consistent with the genetic code.

2. Similarities of the Structures of Amino Acids and Nucleic Acid Bases

2.1. Comparison of amino acid *R* groups with nucleic acid base profiles

We have already proposed stereochemical relationships between amino acids and nucleic acids after finding with CPK space-filling models that the *R* groups of some amino acids could be aligned with the functional groups of nucleic acid bases by intercalation (Hendry and Witham, 1979). The interactions of amino acids with RNA base cavities described in our previous study can also be applied to DNA base cavities. It has since occurred to us that there may be more direct stereochemical relationships between amino acid *R* groups and the purines and pyrimidines of DNA. We have therefore compared different conformations of CPK models of the common L-amino acids to the profiles of thymine (T), cystosine (C), adenine (A) and guanine (G). The α-amino nitrogen of each amino acid was fixed in a position corresponding to the

Diagram I. General structure of an amino acid compared to DNA bases cytosine (C), thymine (T), adenine (A), and guanine (G).



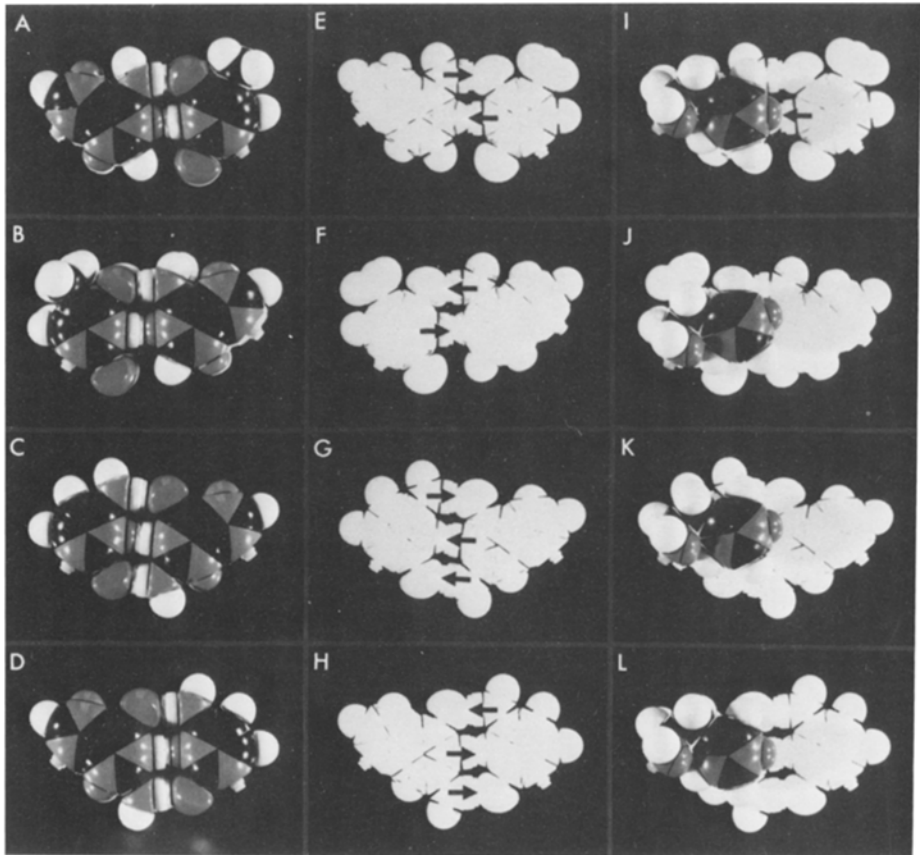


Fig. 1. Proposed structural similarities between CPK models of amino acids and profiles of nucleic acid bases as indicated. Arrows indicate complementary hydrogen bonding sites. Photographs of base pairs are oriented left to right: (A) adeneine-thymine base pair; (B) thymine-adenine base pair; (C) cytosine-guanine base pair; (D) guanine-cytosine base pair; (E) profile of adenine-thymine; (F) profile of thymine-adenine; (G) profile of cytosine-guanine; (H) profile of guanine-cytosine; (I) conformation of L-histidine overlapping profile of adenine; (J) L-histidine as in (I) overlapping profile of thymine; (K) L-histidine as in (I) overlapping profile of cytonsine; (L) L-histidine as in (I) overlapping profile of guanine. See the text for discussion.

9-nitrogen of purines and 1-nitrogen of pyrimidines as shown in the diagram above. The α-CH moiety was also placed in a position approximating that of the 6-CH of pyrimidines and the 8-CH of purines. Amino acids *R* groups were then placed in conformations which would overlap the remainder of the profile of each base as in Diagram I.

Several of the conformations of the amino acids possessed profiles which were similar to those of one or more of the bases (Figures 1 and 2). For example, L-histidine could assume a conformation similar to the profile of A (Figure 1I) so that the proton acceptor at N-3 of the imidazole ring was aligned with the proton acceptor of A at N-1. While the profile of histidine was also similar to the profile of G in the conformation shown (Figure 1L), the N-3 of the imidazole was aligned with the proton donor at N-1 of G. When the same conformation of histidine was compared to the profile of T or C, it was evident that the amino acid *R* group was larger than either of the bases (Figure 1J, K). Although we compared other conformations of the histidine *R* group to the bases in this fashion, we were unable to find any other profiles of the amino acid which were similar to T, C or G. L-arginine could assume a conformation in which each nitrogen of the guanido group was aligned with the N-3, N-1 and the NH_2 groups of G so that the hydrogen bonding donor surface of arginine was congruent with that of the bases (Figure 2A). L-tryptophan and L-glutamic acid could be placed in conformations which were similar to the profiles and the hydrogen bonding character of G and A respectively (Figure 2B and 2C). In the case of amino acids which did not possess potential hydrogen bonding *R* groups, conformations could be constructed which were similar in size to a given base profile, e.g. T for isoleucine, and C for proline (Figure 2D and 2E).

2.2. Comparison of amino acid *R* groups to cavities formed by removal of bases from DNA

Our initial observations suggested that an amino acid might not only be similar in profile to a given nucleic acid base or bases, but that many also possessed stereochemical topologies which could 'fit' the adjacent complementary nucleic acid bases(s) e.g. the potential alignment of hydrogen bonding surfaces of some amino acid conformations with those of the adjacent base (Figure 1, 2A-C). To investigate this possibility further, we substituted amino acids for bases in models of double stranded nucleic acids. Four double stranded DNA models were constructed in an approximate right handed helical conformation and then a single base (T, C, A or G) was removed. While being careful to maintain the approximate stereochemistry of the original DNA models, we placed different conformations of each amino acid in the cavities so that each α-amino nitrogen was covalently bound to C-1 of deoxyribose. The α-carboxy group of the amino acid was allowed to extend into the minor groove in the 5′ direction. Several L-amino acids could be easily accomodated in this fashion without any obvious strain on the model of the double helix, particularly if the amino acid was placed in a conformation similar to the profile of a given base. For example, glutamine could be substituted for A, histidine for A, arginine and tryptophan for G, isoleucine for T and proline for C (Figure 3). In some cases, the size of a given cavity

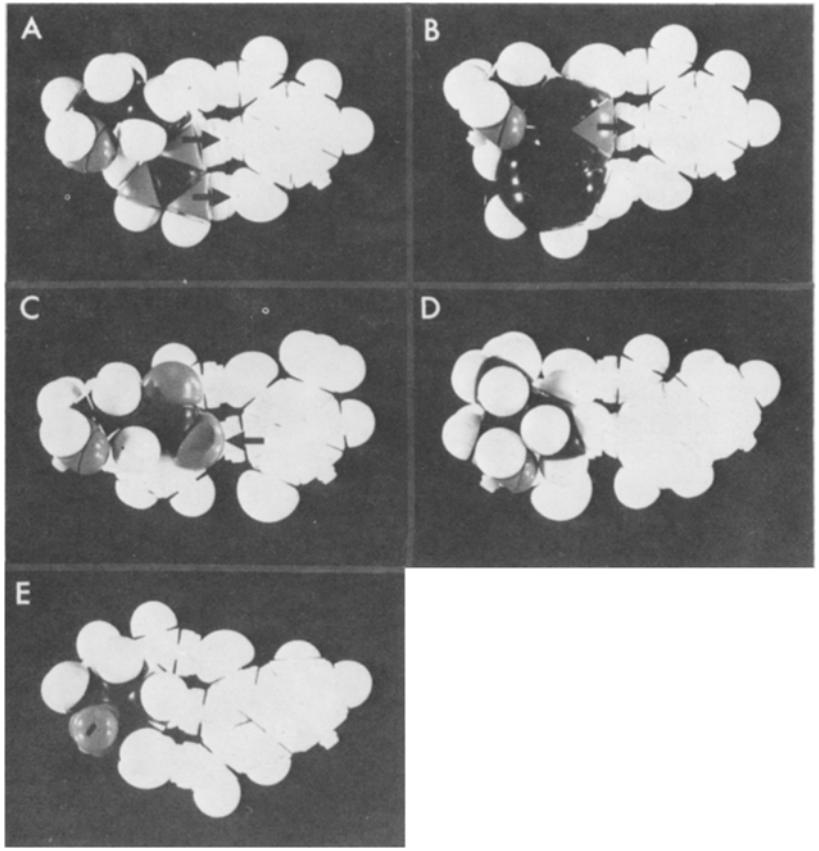


Fig. 2. Proposed structural similarities between CPK models of amino acids and profiles of nucleic acid bases. Arrows indicate complementary hydrogen bonding sites. (A) conformation of L-arginine overlapping profile of G in the G-C base pair; (B) conformation of L-tryptophan overlapping profile of G in G-C base pair; (C) conformation of L-glutamic acid overlapping profile of A in A-T base pair; (D) conformation of L-isoleucine overlapping profile of T in T-A basee pair; (E) conformation of L-proline profile of overlapping C in C-G base pair.

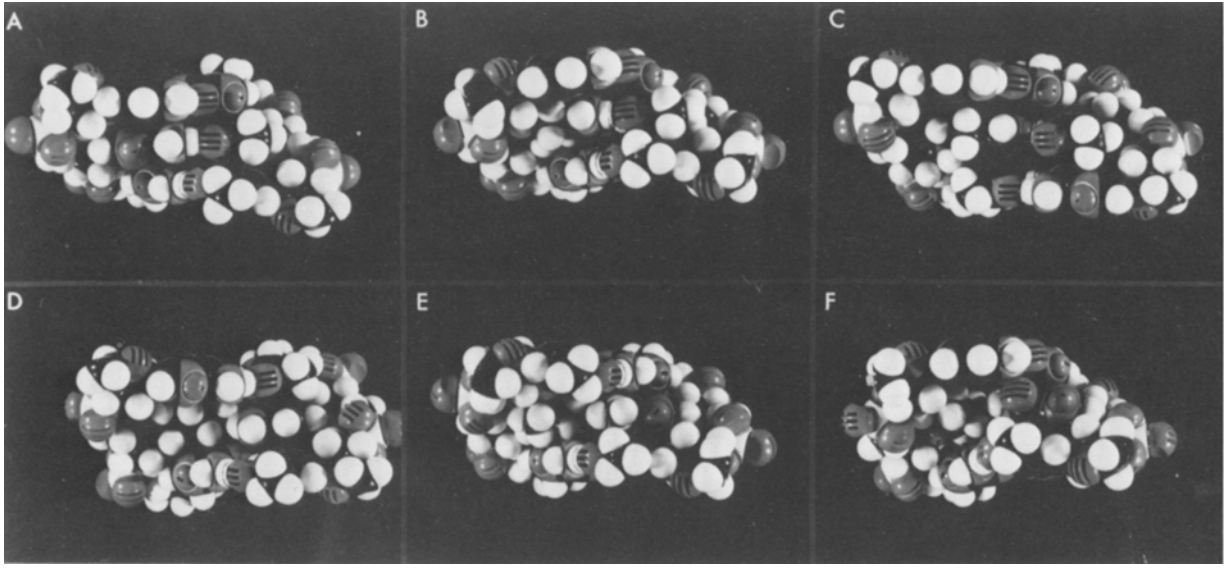


Fig. 3. CPK models of proposed amino amino acid-DNA complexes. All views are of the major groove with the sugar-phosphate backbone from top to bottom 5′-3′ on the left and 3′-5′ on the right. (A) double stranded DNA sequence CG, AT, AT; (B) replacement of L-glutamine in double stranded DNA for A in sequence CG, GLN-T, AT; (C) replacement of L-histidine in double stranded DNA for A in sequence CG. *HIS*-T,,TA; (D) replacement of L-arginine in double stranded DNA for G in sequence AT, *ARG*-C, AT; (E) replacement of L-isoleucine in double stranded DNA for T in sequence TA, *ILEU*-A, AT; (F) replacement of L-proline in double stranded DNA for C in sequence CG, *PRO*-G, AT.

TABLE I

Proposed structural similarities between L-amino acids and nucleic acids. The size of each amino acid is compared to the size of the cacity formed after removal of a base (T, C, A or G) in right handed double helical DNA. > indicates the amino acid *R* group is larger than the cavity; < indicates the *R* group is smaller than the cavity. If an amino acid 'fits' into the cavity, complementarity to the adjacent base is indicated by =. Hydrogen bonding sites on adjacent bases are indicated by letters and numbers which refer to the positions of atoms and functional groups on the nucleic acid structures

L-amino acid	Size compared to DNA base cavity				Complementarity to adjacent base				Positions of hydrogen bonding on adjacent bases	Tentative base substitution
	C	T	A	G	C	T	A	G		
Alanine		<	<	<				=		C
Arginine	>	>			=				N-3, O-2	G
Aspartic acid						=			NH-3	A
Asparagine							=		N-1, NH_2-6	T
								=	NH-1, O-6	C
						=			NH-3	A
Cysteine	>				=				N-3*	G
Glutamic acid	>	>				=			NH-3	A
Glutamine	>	>			=				N-3, NH_2-4	G
						=			NH-3, O-4	A
Histidine	>	>				=			NH-3	A
Isoleucine	>		<	<			=			T
Leucine	>		<	<			=			T
Lysine	>	>			=				O-2	G
						=			O-2 or O-4	A
Methionine	>						=			T
Phenylalanine	>	>								+
Proline			<	<				=		C
Serine			<	<				=	NH-1	C
Threonine			<	<				=	NH-1	C
Tryptophan	>	>			=				N-3	G
Tyrosine	>	>			=				O-2	G
						=			O-2, O-4	A
Valine	>		<	<			=			T

* Direction of hydrogen bonding groups is complementary to adjacent base but distance is longer than normal hydrogen bond.

† Phenylalanine did not possess features which were clearly complementary to the adjacent base.

did not fit any conformation of a given amino acid; in other cases hydrogen bonding groups were not complementary to the base adjacent to the cavity.

Our observations of the similarities of the amino acids to base cavities have been summarized in Table I by listing two factors, size and complementarity. First, we determined whether a given amino acid was roughly too large or too small for a given cavity. An amino acid was too large if it could not fit into the cavity without obviously distorting the nucleic acid components and too small if it could not come into proximity with the adjacent base. If an *R* group was neither too larger or too small for a given cavity, we then considered whether or not the amino acid possessed features which were complementary to the adjacent base. When amino acids contained hydro-

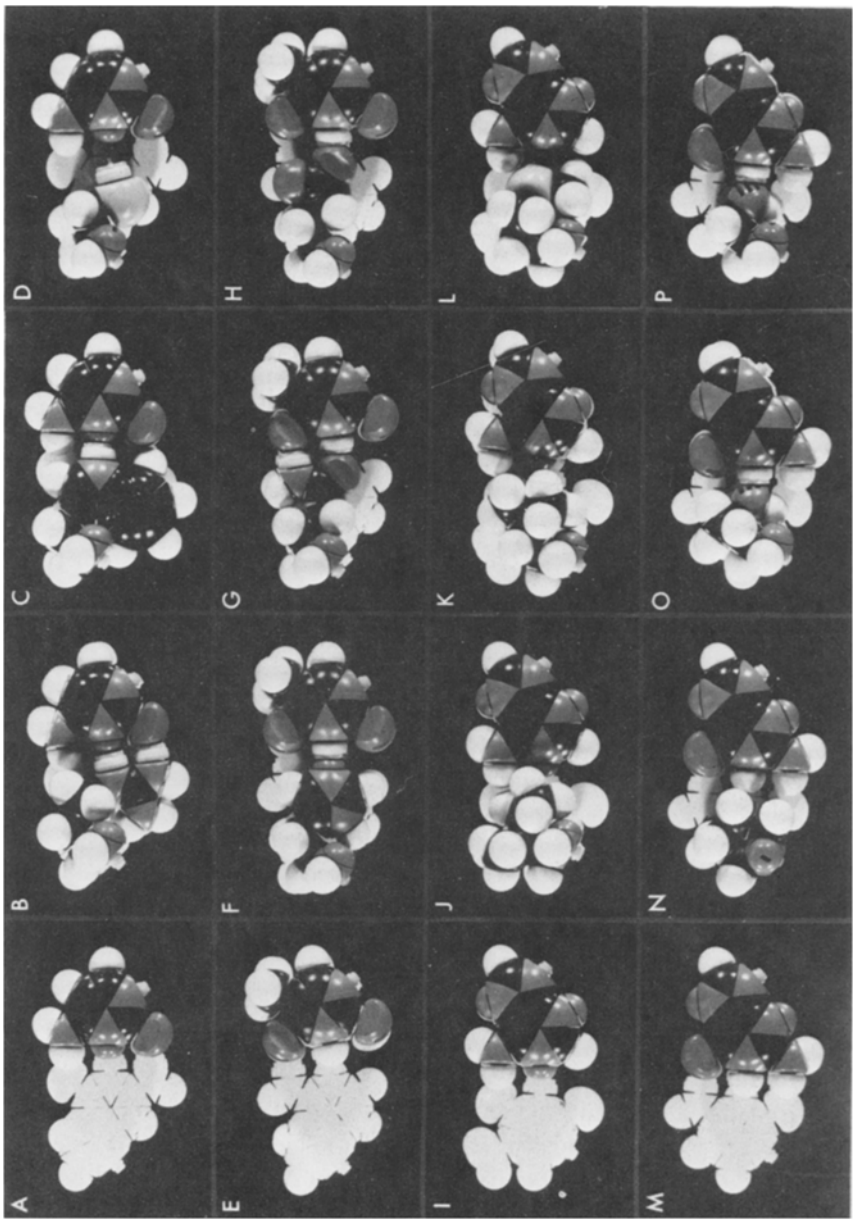


Fig. 4. Proposed structural similarities between CPK models of L-amino acids and nucleic acids. Hypothetical complexes of amino acids and adjacent nucleic acids above profiles of base pairs. (A) cytosine overlapping profile of G-C base pair; (B) arginine-cytosine complex; (C) tryptophan-cytosine complex; (D) cysteine-cytosine complex; (E) thymine overlapping profile of A-T base pair; (F) histidine-thymine complex; (G) glutamine-thymine complex; (H) glutamic acid-thymine complex; (I) adenine overlapping profile of T-A base pair; (J) isoleucine-adenine complex; (K) leucine-adenine complex; (L) methionine-adenine complex; (M) guanine overlapping C-G base pair; (N) proline-guanine complex; (O) threonine-guanine complex; (P) serine-guanine complex.

gen bonding moieties, attempts were made to align the hydrogen bonding surface of the *R* group with complementary groups of the adjacent base. The specific hydrogen bonding groups of the adjacent base(s) which were complementary to a given amino acid are also listed in Table I. These relationships should theoretically reflect C-G and A-T base pairing. When amino acids could be accomodated to a given cavity but lacked hydrogen bonding moieties, *R* groups were considered complementary when a given conformation 'fitted' the surfaces of the adjacent base. Of the twenty common amino acids, fourteen could be considered analagous in structure to a single base. Three amino acids were considered related to two bases and one could be related to three bases.

Figure 4 contains some examples of amino acid conformations which could be accommodated to the size of a given base cavity and also possessed complementary features. Glycine is omitted from Table I; it was not compared to any base since it lacks an *R* group.

Some of the amino acids did not have a clear structural similarity to any of the bases. Although the *R* group of phenylalanine was similar in size to A ro G, it did not appear to be complementary to any adjacent base. Cysteine was closer in size to T than to G or A; G is listed in Table I because the *R* group could adopt a conformation in which the proton of the SH group of the amino acid could be directed toward the proton acceptor at N-3 of the adjacent base C. Such a hydrogen bond, however, would be longer than the normal hydrogen bond distance. As indicated by just these two examples, it should be obvious that we present the relationships listed in Table I as initial approximations which will require further modifications. Leucine, isoleucine and valine might be related to pyrimidines in general rather than just to T because of their approximate similarities in size to both T and C. It is difficult at this point to decide how much significance to attach to the surface complementarity of an amino acid and the adjacent base in the cavity, but such a consideration would favor listing T because leucine (Figure 4K), isoleucine (Figure 4J), and valine 'fit neatly' on the surface of A presented to the cavity.

Our choice of stereochemical constraints no have been somewhat arbitrary, inasmuch as little consideration has been given to: (1) the relative stability of various amino acid conformations; (2) variations of the conformations of the deoxyribose moieties; (3) the possibility that modifications in nucleic acid structure might occur as a result of the insertion of amino acids into the DNA double helix; (4) the relationships of the putative amino acid complexes to neighboring bases; (5) the absolute chirality of the amino acids; (6) hydrophobic interactions of the amino acids in the complexes; (7) the directionality and strength of hydrogen bonds and (8) the relative energetics of the complexes. Also there are other ways of positioning amino acid *R* groups so that the amino acids resemble purines and pyrimidines. For example, we have employed a similar approach with the carboxyl group of amino acids attached via an ester linkage to a 2′ OH of the neighboring 5′ ribose in the base cavity (a bond which is known to occur in t-RNA) rather than linking the α-amino groups to the C-1 of ribose. The relationships of amino acids to base cavities were similar.

3. Assignments of Amino Acids To Nucleic Acid Bases

3.1. Preliminary measurements of amino acid-nucleic acid relationships

In order to assess the structural similarities proposed in Table I, we have attempted to make semiquantitative measurements of these relationships using size and complementarity as parameters. With regard to size, we measured the minimum area of each amino acid which in any *R* group conformation would extend beyond the hydrogen bonding surface of an adjacent base when the amino acid was placed in a cavity in DNA. For complementarity of the *R* groups to the adjacent base, two factors were considered: the ability of the *R* group to form appropriate hydrogen bonds to the adjacent base and how well it would 'fit' neatly with the surfaces of the adjacent base. While neither feature easily lends itself to direct quantitation, we have assumed that: if an amino acid *R* group was not too large for given cavity, we would utilize the area of the *R* group surface which was within a given distance of the adjacent base as a measure of the extent of the potentially complementary surfaces of an amino acid and the adjacent base.

It was necessary to choose fixed conformations of the cavities formed by removal of a base from CPK models of DNA to perform these measurements. When simulating the removal of a base and the replacement with an amino acid, the absolute conformation of deoxyribose was maintained and the width of all *R* group conformations was limited to the approximate width of the DNA cavity. Specific measurements were then made using the following protocol.

(1) Each amino acid was attached covalently via the α-amino group to deoxyribose at C-1. The deoxyribose was kept in the C(3′)-exo conformation such that the positions of the sugar and the attached amino acid approximated that of a base in the *B* conformation of DNA i.e. the α-CH of each *R* group and the α-amino group were positioned approximately at the 6-CH and 1-N of pyrimidines or the 8-CH and 9-N of purines, respectively.

(2) The *R* group was then fitted to the surface of the profile of each base as before so that (1) it did not extend above the surface of a purine or pyrimidine; (2) it was in a plane at an $\sim 156°$ angle (X′) to the deoxyribose ring corresponding to that of a base attached to deoxyribose and did not exceed $\cong 2.25$ Å on either side of the plane.

(3) The minimum area which extended beyond the hydrogen bonding surface of the adjacent base (A), and the maximum area of the amino acid *R* group which was contained within approximately 1.3 Å of the surface of the adjacent base (B) was determined for each amino acid. These measurements, made from photographs of profiles of CPK models of amino acid conformations, are listed in Table II in square Ångstroms. The distances (1.3 Å and 2.25 Å) which we have chosen seem to be reasonable approximations for the potential interactions of an amino acid *R* group with the surface of an adjacent base, as well as for the maximum width of any *R* group which could fit into the cavity.

TABLE II

Preliminary assignments of amino acid *R* groups to cavities in nucleic acids formed by removal of bases (see text)

When an amino acid replaces a base, (A) denotes the area of an *R* group that extends beyond the surface of the adjacent base; (B) denotes the area of an *R* group for which A = 0 that is within 1.3 Å of the adjacent base. The areas given are approximations in square angstroms based upon measurements of photographs of amino acid conformations compared to photographs of the base pairs. Complementary hydrogen bonding *R* groups are indicated by *. Preliminary assignments are in *italics*. See text for discussion.

Base removed in double stranded right handed helical DNA

Amino acid		Cytosine	Thymine	Adenine	Guanine
Alanine	A	*0.0*	0.0	0.0	0.0
	B	*0.2*	0.0	0.0	0.0
Arginine	A	12.0	7.8	0.0	*0.0*
	B	—	—	5.0	*6.8**
Aspartic acid	A	0.0	0.0	*0.0*	0.0
	B	3.8	5.3	*1.2**	1.6
Asparagine	A	*0.0*	*0.0*	*0.0*	0.0
	B	*4.5**	*5.6**	*1.8**	1.1
Cysteine	A	0.9	0.0	0.0	*0.0*
	B	—	2.5	0.9	*0.4**
Glutamic acid	A	2.3	3.1	*0.0*	0.0
	B	—	—	*3.7**	5.3
Glutamine	A	2.7	4.0	*0.0*	*0.0*
	B	—	—	*6.2**	*4.9**
Histidine	A	14.0	8.5	*0.0*	0.0
	B	—	—	*5.8**	6.4
Isoleucine	A	1.4	*0.0*	0.0	0.0
	B	—	*9.4*	0.1	0.1
Leucine	A	2.4	*0.0*	0.0	0.0
	B	—	*14.0*	2.2	2.3
Lysine	A	5.7	3.8	*0.0*	*0.0*
	B	—	—	*5.5**	*4.5**
Methionine	A	3.3	*0.0*	0.0	0.0
	B	—	*6.5*	4.7	5.0
Phenylalanine	A	9.5	5.4	*0.0*	0.0
	B	—	—	*5.0*	4.5
Proline	A	*0.0*	0.0	0.0	0.0
	B	*2.9*	1.5	0.0	0.0
Serine	A	*0.0*	0.0	0.0	0.0
	B	*1.7**	1.1	0.0	0.3
Threonine	A	*0.0*	0.0	0.0	0.0
	B	*1.6**	1.4	0.0	0.0
Tryptophan	A	16.3	9.7	0.0	*0.0*
	B	—	—	4.9	*6.7**
Tyrosine	A	14.0	8.5	*0.0*	*0.0*
	B	—	—	*5.8**	*6.4**
Valine	A	0.3	*0.0*	0.0	0.0
	B	—	*2.4*	0.0	0.0

TABLE III

Comparison of amino acid *R* groups to nucleic acid bases. Results from scheme below. Comparison of amino acids to bases replaced (A) cytosine; (B) thymine; (C) adenine; (D) guanine.

Question 1 – Does the amino acid *R* group not extend beyond the surface of the adjacent base?

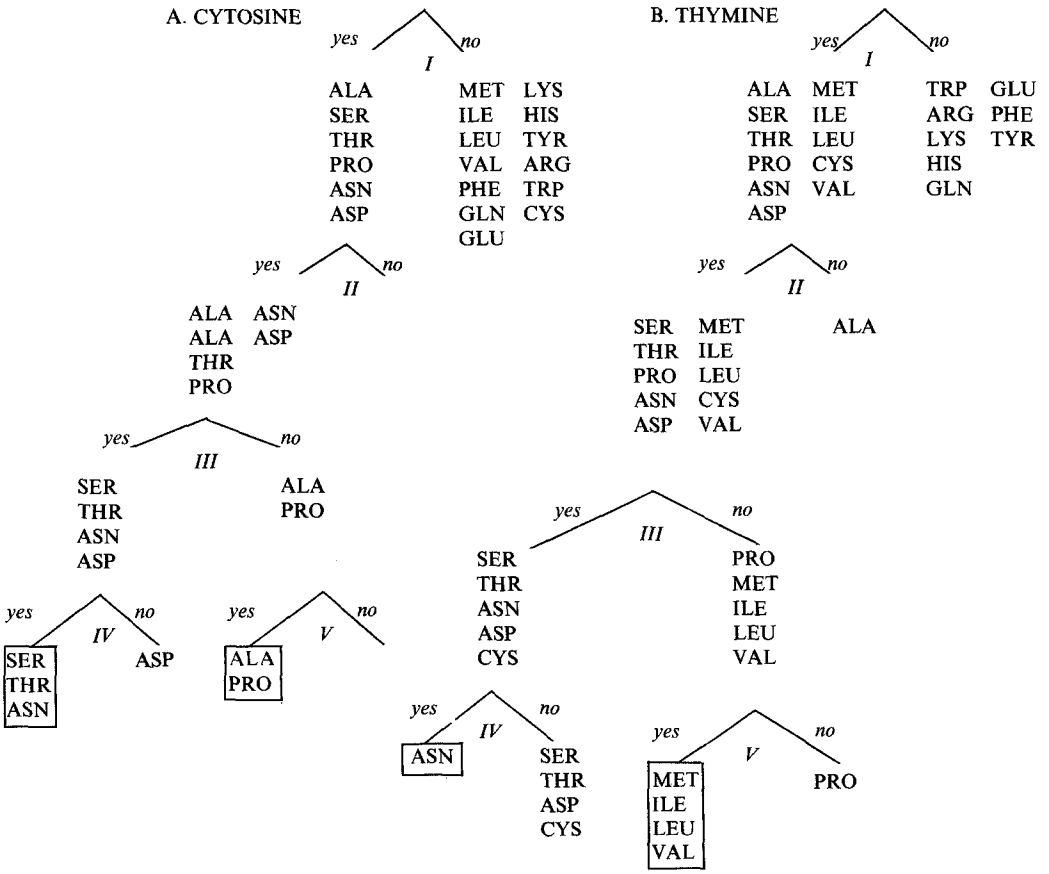
Question 2 – If 1 is yes, is the amino acid within 1.3 angstroms of the adjacent base?

Question 3 – If 2 is yes, does the amino acid possess hydrogen bonding functional groups?

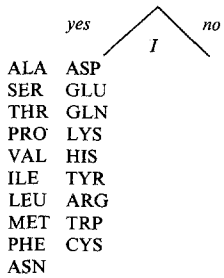
Question 4 – If 3 is yes, can complementary hydrogen bonds form between the amino acid *R* group and the adjacent base?

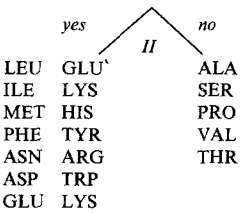
Question 5 – If 3 is no, is the surface area of the *R* group which is 1.3 angstroms from the adjacent base greater than the surface area of the *R* group for other adjacent bases? (Amino acids whose *R* groups extend beyond the adjacent base are excluded).

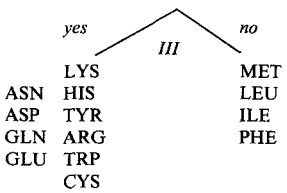
The amino acids which are in boxes were assigned to those bases

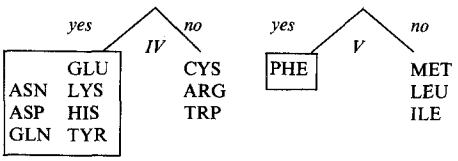


C. ADENINE

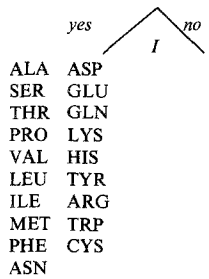


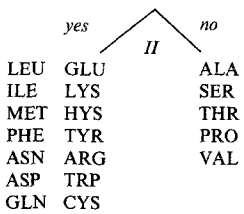


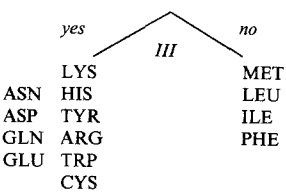


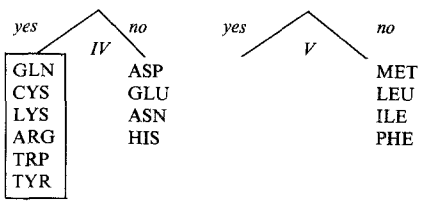


D. GUANINE









3.2. Assignments of amino acids to DNA bases

Before making the assignments in Table II, we excluded from consideration any amino acids which had *R* groups that either extended beyond the hydrogen bonding surface of the adjacent base in the cavity, (A > O) or had *R* groups which were not within approximately 1.3 Å of the surface of the adjacent base (B = O). The results are consistent with the similarities in structure listed in Table I. In the case of amino acids possessing hydrogen bonding functional groups, those bases for which the surface complementarity parameter *B* was greater than zero (B > O) but which were not complementary to the adjacent base were not assigned. For example, when L-histidine was compared to the cavity formed by removal of either A or G, B was greater than zero. Histidine was then assigned to A since the proton acceptor at N-3 of the imidazole ring could be hydrogen bonded to the adjacent base T at N-1. G was not assigned because the proton acceptor at N-1 of C prohibited such a hydrogen bond. Amino acids which do not possess hydrogen bonding functional groups were assigned to that base which had the largest value for surface complementarity (B) and hence would have the greatest potential contact with the adjacent base. A scheme is presented in Table III which describes how the amino acids were assigned to each base using these criteria.

4. Relationships of Amino Acids to the Genetic Code

4.1. Comparison of amino acid assignments to bases in their codons

During protein synthesis amino acids are coded for by triplets of nucleic acid bases known as codons. Consequently, it seemed reasonable to compare those codon assignments which comprise the genetic code to our assignments of amino acids to nucleic acid bases. There is a good correlation between our assignments and the second base in the codons of many amino acids. They are underlined in Table IV. Of the fifteen amino acids which we could assign to a single base, the base assigned is the second base in forty-five of the fifty-one codons attributed to those amino acids. The three amino acids for which two bases were chosen, and asparagine for which three bases were chosen, have been assigned bases which were consistent with the second base of their codons. The assignments of the amino acids which were inconsistent with the second base of their codons were phenylalanine and serine. Serine (assigned to C) has two different bases (C and G), which occur at the second position in its codons. Phenylalanine has U as its second codon base whereas A was assigned.

4.2. Correlation with previous assignments of amino acids to their codons

That some amino acids can be correlated in structure to the second base in their codons is interesting in light of our previous suggestion (Hendry and Witham, 1979) that amino acids could be recognized stereochemically by certain bases of their codons and anticodons. When the α-amino groups of L-amino acids were attached to the phosphate oxygens of CPK models of double stranded RNA and the *R* groups

TABLE IV

Relationships of assignments of L-amino acids to nucleic acids to their codons

amino acid	Assignments from comparison to DNA base cavities[+]	Assignments from[++] intercalation of amino acids between RNA bases[49]	Combined[++] assignments (see text)	Codons
Alanine	C	NA*	UC,CC,AC,GC	GCU,GCC,GCA,GCG
Arginine	G	UG,CA,CG,AG	UG,CG,AG	CGU,CGC,CGA,CGG,AGG,AGA
Aspartic	A	UU,UC,UA,CU,CC,AU,AA,AG,GC,GA	UA,AA,GA	GAU,GAC
Asparagine	C,U,A	UU,UG,CA,AA	UU,CA,AA	AAU,AAC
Cysteine	G	UU,UG	UG	UGU,UGC
Glutamic acid	A	AA,AG,GA,GG	AA,GA	GAA,GAG
Glutamine	A,G	CA	CA	CAA,CAG
Glycine	NA*	GG	GG	GGU,GGC,GGA,GGG
Histidine	A	CA	CA	CAU,CAC
Isoleucine	U	UU,UC,UA,AU,AC,UG	UU,AU	AUU,AUC,AUA
Leucine	U	UU,CU	UU,CU	UUA,UUG,CUU,CUC,CUA,CUG
Lysine	A,G	AA	AA	AAA,AAG
Methionine	U	UU,AU	AU,UU	AUG
Phenylalanine	A	UU	NA*	UUU,UUC
Proline	C	CC	CC	CCU,CCC,CCA,CCG
Serine	C	NA*	UC,CC,AC,GC	UCU,UCC,UCA,UCG,AGU,AGC
Threonine	C	UU,UC,CU,CC,CA,AC	CC,AC,UC	ACU,ACC,ACA,ACG
Tryptophan	G	UG	UG	UGG
Tyrosine	A,G	UU,UA,UG	UA,UG	UAU,UAC
Valine	U	NA*	UU,CU,AU,GU	GUU,GUC,GUA,GUG

* Indicates no assignment made.

[+] Underlined bases are consistent with the second base of their codon(s).

[++] Underlined bases are consistent with the first two bases of their codons (U = T).

were allowed to intercalate between two consecutive bases, certain conformations of several amino acids 'fit' into cavities formed between the bases and were able to form complementary complexes with the surfaces of both the 3′ and 5′ adjacent bases of the opposite strand. Amino acids were assigned bases based upon their ability to 'fit' between bases of one strand and form complementary complexes to both the 3′ and 5′ bases of the opposite strand. These assignments given in Table IV correlate well with the intercalation of the amino acids between the first two bases of their codons in the 5′ to 3′ direction and complementary pairing of the amino acid *R* groups with the adjacent 3′ and 5′ bases of their anticodons. If our current assignments by base substitution of amino acids to nucleic acid bases is considered as indicative of the second base of a triplet code, and thus is used to refine our previous intercalation assignments, we end up with a list of combined assignments (Table IV) which is very similar to a list of the first two bases of the genetic code.

5. Conclusions

Although any stereochemical model of the relationships of amino acids and nucleic acids may be biased by the fact that the genetic code is already known, our findings suggest that structural relationships between nucleic acid bases and the common L-amino acids may exist. Moreover, the similarities in size and topology of various *R* groups and DNA bases, in particular those features of some *R* groups which form analogous complementary pairs with certain bases, raises the possibility that some amino acid *R* groups could be stereochemically recognized by nucleic acids and vice versa. In this regard, we would agree with Woese (1967) who contended '...in my opinion, it is difficult not to conclude that codon-amino acid pairing played a major role in the shaping of the genetic code'. It seems reasonable therefore to revive the hypothesis that the synthesis of biologically essential polymers e.g. enzymes, receptors and immunoglobulins, principally from only twenty structures out of an almost infinite number of possible structural units is a result of unique structural relationships between amino acids and nucleic acids.

It has not been our intention in this paper to postulate the exact stereochemistry of amino acid-nucleic acid interactions. Not only has our construction of CPK models failed to consider all possible configurations of the two ligands, but there must have been many errors inherent in our constructing three dimensional models and in our two dimensional measurements of area. Our assignments moreover are not yet fully compatible with the triplet code for amino acids which was deduced from peptide synthesis *in vitro* with synthetic oligonucleotides used as templates. Some of our selections are not consistent with the second base of known codons (Table IV). For example, asparagine can replace C and T in addition to A, the second base of its codon. Replacement of T or C could even be considered better assignments since they can form two hydrogen bonds whereas when replacing A only one hydrogen bond can be formed. Serine has two second bases in its codons, C and G; we assigned only C. Serine is inappropriate in size for replacing G (B = O) even though a hydrogen bond could be envisioned between the hydroxyl and N-3 of C, the adjacent base.

Phenylalanine does not adequately replace any base, although it was assigned A based upon size alone. We cannot explain this glaring exception.

Clearly, the ability of amino acids and nucleic acids to form complementary surfaces is not described adequately by our preliminary measurements made from CPK models. Experimental approaches which are successful in quantitating complementarity would be helpful and might satisfy the requirement of the 'knobby hydrophobic surfaces' (Melcher, 1974) which Crick (1968) has deemed important for the development of a stereochemical rationale for the genetic code. Consideration should be given to surface complementarity of amino acid *R* groups to adjacent bases in DNA cavities without the dimensions fixed. The effect of alternative types of base pairing (e.g. Hogstein) as well as potential interactions of *R* groups with tautomeric forms of the bases or with rare bases such as occur in t-RNA will also need to be appraised. In this regard, Seeman *et al.* (1976) in postulating a mechanism for sequence-specific recognition of double helical nucleic acids by proteins used a different approach form ours but reached similar conclusions, i.e., that amino acid side chain-nucleic acid pairing is an important factor in the recognition process.

Our observations are consistent with primacy of the second base of the genetic code (codon or anticodon) for which there is some evidence. Alff-Steinberger (1969) concluded that 'the second base position in the codon plays the largest role in determining the properties of the amino acids' after considering a computer generated code which appraised molecular weight, polar requirement, number of dissociating groups, pKl, isoelectric points, and the α-helix forming ability of amino acids. Woese (1967) had stated a similar view based on among other things error rates in translation as being 100:10:1 when considering the III:I:II base positions in the codon. Similar conclusions were reached by Volkenstein (1966) and Dickerson (1970) based upon studies of protein structure and by Zhdanov (1974) who correlated oxidation state of the amino-acid *R* groups with the second codon base. Nagvary and Fendler (1974) and Weber and Lacey (1978) correlated amino acid properties and in particular polarity of the second base in the anticodon with the relative polarity of the amino acids. Recently, Jungck (1978) also showed that the polarity and bulkiness of the amino acid can be correlated to anticodon bases. While it is not possible to discuss and fully review previous work on the genetic code in this communication, much of the published data when considered in view of our models appears to be in agreement with our findings. For example, we would predict from our model complexes that the polarity of the second base of the anticodon may be an important factor in the complementarity to an amino acid.

We are currently attempting to provide better descriptions of the DNA – amino acid complexes using a combination of techniques including computer manipulation of available X-ray structures coupled with an examination of CPK and the more accurate Kendrew framework models. In this regard, we maintain the view that the principle of chemical complementarity which arose from the classical experiments of Chargaff (1950) with nucleic acid bases, may have analogies in the evolution of structural constraints on amino acids coded for in the construction of protein.

Acknowledgements

The photographs contained in this manuscript were kindly provided by Mr Peter Ide, of the Medical College of Georgia. We also wish to acknowledge the assistance of R. Anthony Hicks, Jannie Jones and of Cheryl Lewis in preparation of the manuscript. We thank Professor Carl Woese, Richard Feldmann, Professor John Reissner. Dr Ken Craig, Professor Rob Weinstein and Professor George Gokel for valuable discussions and suggestions regarding the ideas presented herein. We also wish to acknowledge the contributions of Professor Francis H. Witham who provided much encouragement and insight in the genesis of the general theoretical framework presented.

References

* Alff-Steinberger, C.: 1969, *Proc. Nat. Acad. Sci. (USA)* **64,** 584.
* Batchinsky, A. G., and Ratne, V. A.: 1976, *Origins of Life* **7,** 229.
* Brack, A., and Orgel, L.: 1975, *Nature* **256,** 383.
* Calvin, M.: 1973, *Int. J. Polymeric Mater.* **2,** 291.
Chargaff, E.: 1950, *Experientia* **6,** 201.
* Chernavski, D. S., and Chernavskaya, N. M.: 1975, *J. Theor. Biol.* **50,** 13.
Crick, F. H.: 1968, *J. Mol. Biol.* **38,** 367.
Crothers, D. M., Steno, T., and Soel, D. G.: 1972, *Proc. Nat. Acad. Sci. (USA)* **69,** 3063.
* Danckwerts, H. J., and Neubert, D.: 1975, *J. Mol. Evol.* **5,** 327.
Dickerson, R. E.: 1971, *J. Mol. Biol.* **57,** 1.
* Dunnill, P.: 1966, *Nature* **210,** 1267.
Gamow, G.: 1954, *Nature* **173,** 318.
* Goldberg, A. L., and Wittes, R. E.: 1966, *Science* **153,** 420.
* Grantham, R.: 1974, *Science* **185,** 862.
* Hartman, H.: 1975, *Origins of Life* **6,** 423.
Hendry, L. B., and Witham, F. H.: 1979, *Persp. Biol. Med.* **22,** 333.
Hendry, L. B., Witham, F. H., and Chapman, O. L.: 1977, *Persp. Biol. Med.* **21,** 120.
* Hopfield, J. J.: 1978, *Proc. Nat. Acad. Sci. (USA)* **75** (9), 4334.
* Ishigami, M., and Nagano, K.: 1975, *Origins of Life* **6,** 551.
* Jorre, R. P., and Curnow, R. N.: 1975, *Biochimie* **57,** 1147.
* Jukes, T. H., Holmquist, R., and Moise, H.: 1975, *Science* **189,** 50.
* Jungck, J. R.: 1978, *J. Mol. Evol.* **11,** 211.
* Lacey, J. C., Weber, A. L., and White, W. E.: 1975, *Origins of Life* **6,** 273.
* Lacey, J. C., and Pruitt, K. M.: 1969, *Nature* **223,** 799.
* Lesk, A. M.: 1970, *J. Theor. Biol.* **27,** 171.
Melcher, G.: 1974, *J. Mol. Evol.* **3,** 121.
Mikelsaar, H. N.: 1975, *J. Theor. Biol.* **50,** 203.
* Nagyvary, J., and Fendler, J. H.: 1974, *Origins of Life* **5,** 357.
* Nakashima, T., and Fox, S. W.: 1972, *Proc. Nat. Acad. Sci. (USA)* **69** (1), 106.
* Orgel, L. E.: 1968, *J. Mol. Biol.* **38,** 381.
* Pelc, S. R., and Welton, M. G. E.: 1966, *Nature* **209,** 868.
* Ralph, R. K.: 1968, *Biochem. Biophys. Res. Comm.* **33** (2), 213.
* Ratner, V. A., and Batchinsky, A. G.: 1976, *Origins of Life* **7,** 225.
* Rendell, M. S.: 1972, *J. Am. Chem. Soc.* **94,** 121.
* Rendell, M. S., Harlos, J. P., and Rein, R.: 1971, *Biopolymers* **10,** 2083.
* Salemme, F. R., Miller, M. D., and Jordon, S. R.: 1977, *Proc. Nat. Acad. Sci. (USA)* **74,** 2820.
* Saxinger, C., and Ponnamperuma, C.: 1971, *J. Mol. Evol.* **1,** 63.
* Saxinger, C., and Ponnamperuma, C.: 1974, *Origins of Life* **5,** 189.
Seeman, N. C., Rosenberg, J. M., and Rich, A.: 1976, *Proc. Nat. Acad. Sci. (USA)* **73** (3), 804.

* Thomas, B. R.: 1970, *Biochem. Biophys. Res. Comm.* **40** (6), 1289.
* Volkenstein, M. V.: 1966, *Biochim. Biophys. Acta* **119,** 421.
* Weber, A. L., and Lacey, J. C.: 1978, *J. Mol. Evol.* **11,** 199.
* Welton, M. G. E., and Pelc, S. R.: 1966, *Nature* **209,** 870.
* Wetzel, R.: 1978, *Origins of Life* **9,** 39.
Witham, F. H., Hendry, L. B., and Chapman, O. L.; 1978, *Origins of Life* **9,** 7.
* Woese, C. R.: 1961, *Biochem. Biophys. Res. Comm* **5** (2), 88.
* Woese, C. R.: 1965a, *Proc. Nat. Acad. Sci. (USA)* **54,** 71.
* Woese, C. R.: 1965b, *Proc. Nat. Acad. Sci. (USA)* **54,** 1546.
* Woese, C. R.: 1967, *The Genetic Code: The Molecular Basis for Genetic Expression,* Harper and Row, New York.
* Woese, C. R.: 1968, *Proc. Nat. Acad. Sci.* **59,** 110.
* Woese, C. R.: 1970, *Bioscience* **20** (8), 471.
* Wong, J. T.: 1975, *Proc. Nat. Acad. Sci. (USA)* **72** (5), 1909.
* Yeas, M.: 1969, *Frontiers of Biology* **12,** North Holland, Amsterdam.
Zhdanov, Y. A.: 1974, *Doklady Akademii Nauk SSSR* **217,** 456.