# A mechanism for the association of amino acids with their codons and the origin of the genetic code

**Shelley D. Copley*†, Eric Smith‡, and Harold J. Morowitz§**

*Cooperative Institute for Research in Environmental Sciences, Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309; ‡Santa Fe Institute, Santa Fe, NM 87501; and §Krasnow Institute, George Mason University, Fairfax, VA 22030

Communicated by Gregory A. Petsko, Brandeis University, Waltham, MA, February 7, 2005 (received for review December 13, 2004)

**The genetic code has certain regularities that have resisted mechanistic interpretation. These include strong correlations between the first base of codons and the precursor from which the encoded amino acid is synthesized and between the second base of codons and the hydrophobicity of the encoded amino acid. These regularities are even more striking in a projection of the modern code onto a simpler code consisting of doublet codons encoding a set of simple amino acids. These regularities can be explained if, before the emergence of macromolecules, simple amino acids were synthesized in covalent complexes of dinucleotides with α-keto acids originating from the reductive tricarboxylic acid cycle or reductive acetate pathway. The bases and phosphates of the dinucleotide are proposed to have enhanced the rates of synthetic reactions leading to amino acids in a small-molecule reaction network that preceded the RNA translation apparatus but created an association between amino acids and the first two bases of their codons that was retained when translation emerged later in evolution.**

catalysis | origin of life

The genetic code has many regularities (1), of which only a subset have explanations in terms of tRNA function (2) or robustness against deleterious effects of mutation (3, 4) or errors in translation (3, 5). There is a strong correlation between the first bases of codons and the biosynthetic pathways of the amino acids they encode (1, 6). Codons beginning with C, A, and U encode amino acids synthesized from α-ketoglutarate (α-KG), oxaloacetate (OAA), and pyruvate, respectively.¶ These correlations are especially striking in light of the structural diversity of amino acids whose codons share a first base. For example, codons for Glu and Pro both begin with C, and those for Cys and Leu begin with U. Codons beginning with G encode amino acids that can be formed by direct reductive amination of a simple α-keto acid. These include glycine, alanine, aspartate, and glutamate, which can be formed by reductive amination of glyoxalate, pyruvate, OAA, and α-KG, respectively. There is also a long-recognized relationship between the hydrophobicity of the amino acid and the second base of its codon (1). Codons having U as the second base are associated with the most hydrophobic amino acids, and those having A as the second base are associated with the most hydrophilic amino acids.

We suggest that both correlations can be explained if, before the emergence of macromolecules, simple amino acids were synthesized from α-keto acid precursors covalently attached to dinucleotides that catalyzed the reactions required to synthesize specific amino acids (see Fig. 1). This is a significant departure from previous theories attempting to explain the regularities in the genetic code (3). The "stereochemical" hypothesis suggests that binding interactions between amino acids and their codons or anticodons dictated the structure of the genetic code (7–10). The "coevolution" hypothesis (6) suggests that the original genetic code specified a small number of simple amino acids, and that, as more complex amino acids were synthesized from these precursors, some codons that initially encoded a precursor were ceded to its more complex products. Finally, the genetic code has been proposed to be simply a "frozen accident" (11).

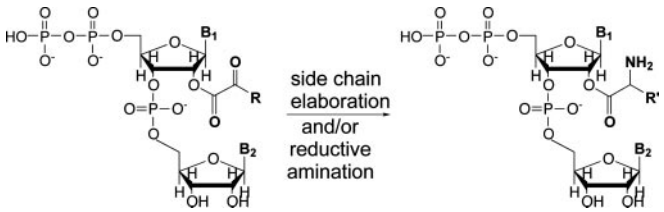


**Fig. 1.** Model for synthesis of amino acids from α-keto acid precursors covalently attached to dinucleotides. The dinucleotide that is capable of catalyzing synthesis of a particular amino acid is proposed to contain the first two bases of the codon specifying that amino acid.

Recent analysis suggests that the reductive tricarboxylic acid cycle could serve as a network-autocatalytic self-sufficient source for simple α-keto acids, including glyoxalate, pyruvate, OAA, and α-KG, as well as the carbon backbones of sugars and nucleobases (12). α-Keto acids can also be generated from the reductive acetyl CoA pathway (13). Most simple amino acids can be reached from an α-keto acid precursor by a small number of relatively simple chemical transformations, and the synthetic pathway that will be followed is determined within the first three steps. We propose that the positions of functional groups in a dinucleotide–α-keto acid complex determine what reactions can be effectively catalyzed for a given α-keto acid. An example of a series of reactions leading from α-KG to five amino acids, each attached to the first two bases of its codon, is shown in Fig. 2, which can be regarded as a "decision tree" in which the nature of the bases in the dinucleotide determines which types of reactions occur. The pathways proposed follow closely those in extant organisms (14), differing primarily in the timing of the reductive amination leading to the final amino acid. The motivation for this approach is that modern biosynthetic pathways likely emerged by gradual acquisition of enzymes capable of catalyzing reactions that had previously occurred in the absence of macromolecular catalysts. Thus, modern pathways are "metabolic fossils" that provide insight into prebiotic synthetic pathways, although some refinements and permutations are expected to have occurred.

The proposition that nucleotides might provide catalytic assistance during synthesis of amino acids is plausible. Catalytic RNAs are capable of catalyzing a wide range of chemical transformations (15–21). Because the amino acid building blocks of proteins can catalyze reactions such as aldol condensation (22, 23), it is reasonable to expect that the nucleotide building blocks of RNAs may also have catalytic abilities. Dinucleotides might catalyze reactions required for synthesis of amino acids by (*i*) orientation and polarization of reactants by hydrogen bonding interactions, (*ii*) use of

Abbreviations: α-KG, α-ketoglutarate; OAA, oxaloacetate.

†To whom correspondence should be addressed. E-mail: shelley@cires.colorado.edu.

¶Correlations between the first bases of codons and biosynthetic pathways were first pointed out by Wong (6) and later by Taylor and Coates (1). However, the earlier work focused on Glu, Asp, and Ala as the synthetic precursors rather than the α-keto acids α-KG, OAA, and pyr.





www.pnas.org/cgi/doi/10.1073/pnas.0501049102

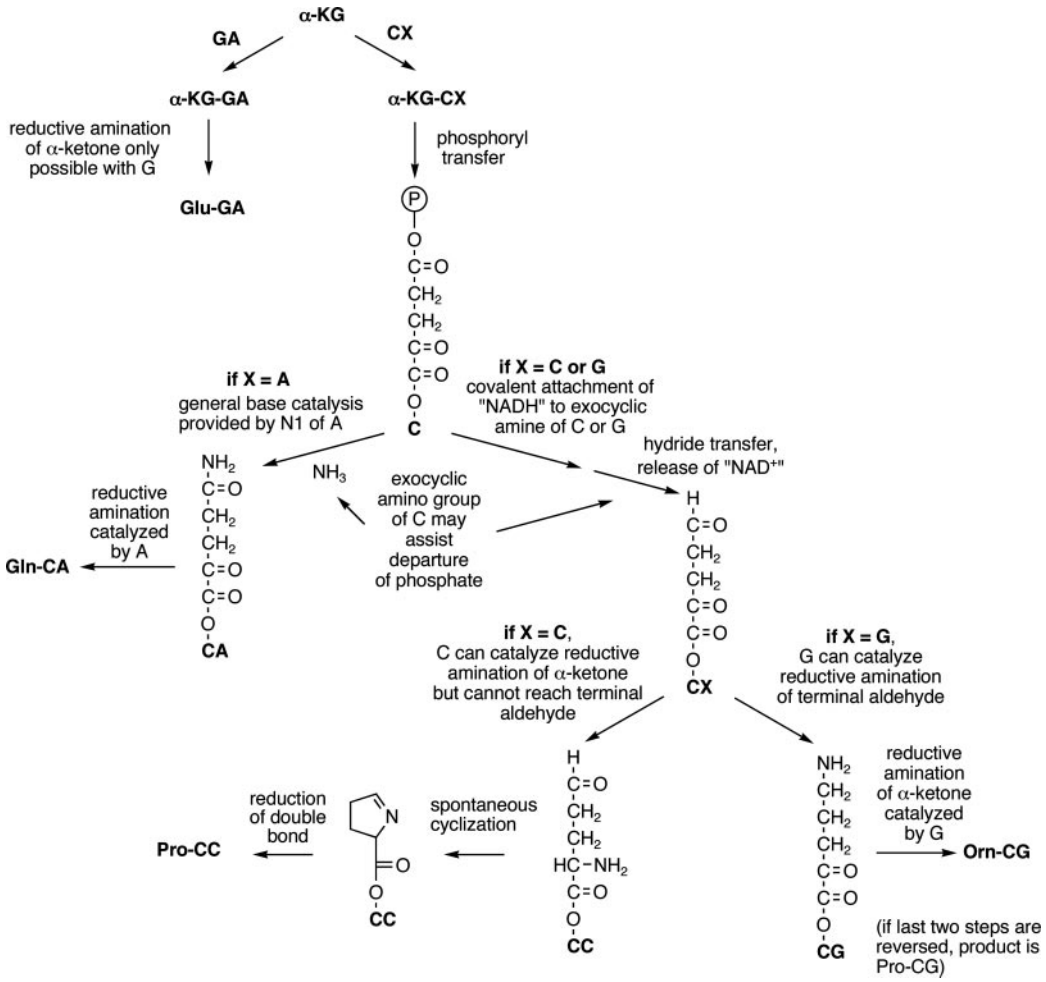


**Fig. 2.** Decision tree showing how attachment of α-KG to a given dinucleotide might influence the chemical reactions that can be facilitated. Reactions with AX or UX are postulated to lead to dead-end products and are consequently not shown.

functional groups as nucleophiles or general bases, (*iii*) attachment of cofactors such as NADH or prebiotic equivalents, (*iv*) use of phosphate groups as general acid or general base catalysts, (*v*) use of $Mg^{2+}$ ions coordinated to phosphate groups as Lewis acid catalysts, and (*vi*) use of $Mg^{2+}$-coordinated hydroxide ions as nucleophiles or general base catalysts. Many of these strategies are used by ribozymes (24–27). Although the level of catalysis achieved by a dinucleotide would likely be modest relative to those achieved by macromolecular catalysts, simply tethering a catalytic group to a molecule can result in enormous rate accelerations. For example, hydrolysis of succinanilic acid occurs $\approx 10^5$-fold faster than hydrolysis of acetanilide due to assistance from the intramolecular carboxylate. Tetramethyl succinanilic acid undergoes hydrolysis 1,200-fold faster than succinanilic acid due to restriction of conformational mobility by the methyl groups on the carbon backbone (28). Another striking example is provided by studies showing that replacement of the 2′-hydroxyl of the tRNA in the P-site of the ribosome with either H or F results in a $10^6$-fold decrease in the rate of peptidyl transfer. The magnitude of this effect suggests that the 2′-hydroxyl is involved in catalysis, either as a general acid or general base or as a coordination site for a catalytically essential metal ion (29). These and other examples illustrate the potential for rate enhancement by intramolecular catalytic groups and restriction of conformational freedom in the reacting centers. We note, however, that dramatic rate enhancements are not necessary. Even modest rate enhancements in the context of a network of chemical reactions can, over time, result in the predominance of certain pathways and components.

Below we present, in two stages, a chemically plausible mechanism for the association of amino acids with the first two bases of their codons. First, we project the modern genetic code onto a simpler doublet code with striking features that are likely to have arisen before the emergence of macromolecular biochemistry and that provide strong clues to the origin of the genetic code. Second, we develop a model suggesting that covalent attachment of an amino acid to a dinucleotide could lead to preferential synthesis of specific amino acids due to catalysis of certain chemical transformations by the dinucleotide.

## Mapping of the Modern Genetic Code onto a Simpler Doublet Code

The 20 amino acids encoded by the modern code do not saturate the bound for the number of amino acids that can be unambiguously encoded by three bases (64) because of the degeneracy of the code. The structurally and synthetically complex amino acids (Trp, Tyr, Phe, Lys, His, and Met) are distinguished from simpler amino acids only in the third position of their codons and are likely to have been added to the code relatively late (30). The 14 remaining amino acids can be specified by using a doublet code and, in fact, these amino acids nearly saturate the bound for the number of amino acids that can be unambiguously encoded by two bases (16). These properties are consistent with later addition of a third base to an established two-base association. They are also consistent with an emergence

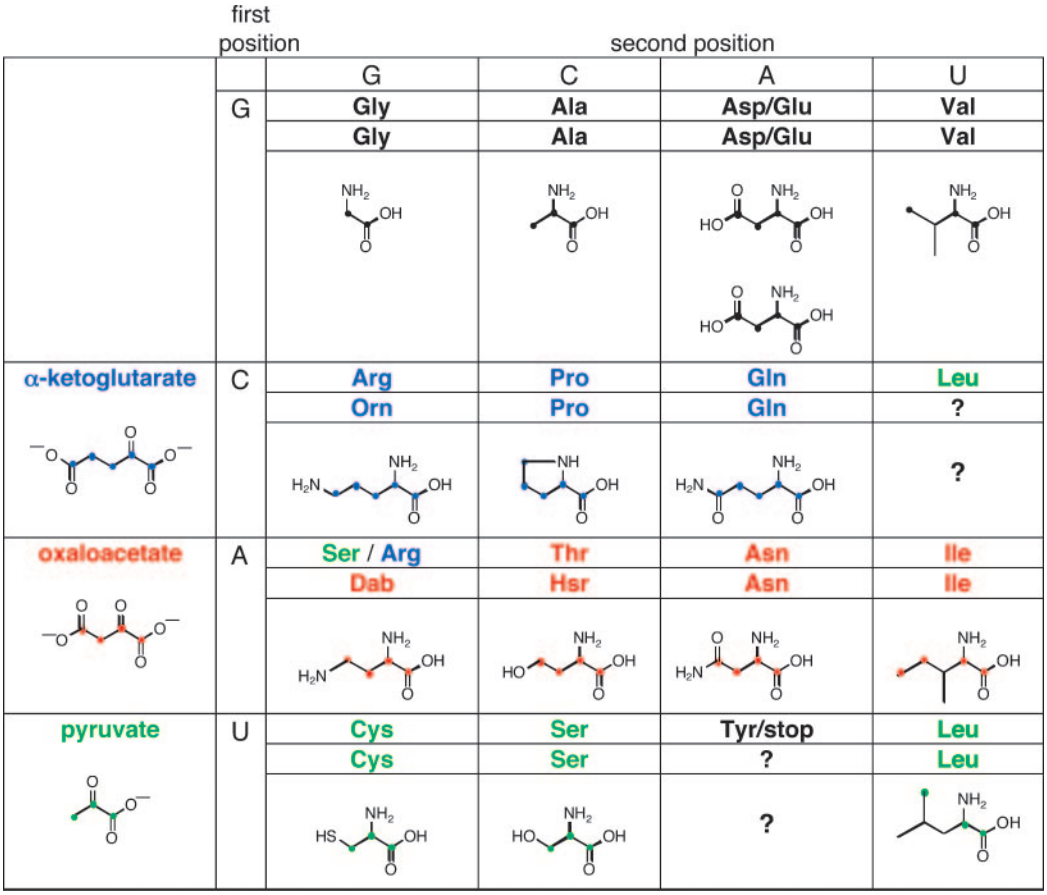


**Fig. 3.** A simplified doublet genetic code produced from the modern genetic code as described in the text. The first line indicates amino acids chosen by application of majority rule, with ties broken in favor of the simplest amino acids. The second line indicates simplifications achieved using simpler precursors of the more complicated amino acids and, in some cases of likely codon reassignments, postulating possible amino acids formed by pathways that are synthetically homologous to those for established amino acids. The third line shows the structures of these amino acids. Black (with the exception of Tyr) indicates amino acids produced by direct reductive amination of α-keto acids; blue, red, and green indicate amino acids synthesized from α-KG, OAA, and pyruvate, respectively. Colored dots indicate carbon atoms derived from the indicated precursors.

of the reading frame and tRNA concurrent with or later than the emergence of the triplet code. We do not seek to explain the full code, the mechanism for addition of the complex amino acids to the code, or the emergence of translation here, although the arguments we provide are a consistent foundation on which to pursue these questions later.

We begin by projecting the modern code onto a simpler doublet code. Many authors have suggested that a simpler code preceded the modern code, but the reasoning and predicted codes have been varied and not closely related to what we will describe (31–35). In assigning amino acids to doublet codons, we have chosen the amino acid assigned to the majority of triplet codons in which the first two bases are common, with ties broken in favor of the simpler amino acid (upper lines of Fig. 3). We choose not to resolve the ambiguity between Asp and Glu for GA, because these amino acids are close structural homologues and structures capable of discriminating between them might have arisen only at a higher level of complexity.

The second stage of mapping onto a simpler code (second lines in Fig. 3) is accomplished by identifying cases for which a simpler precursor might have occupied a position later taken by a more complex amino acid and positions in which the modern assignments are likely to be due to codon reassignments (36, 37). We have assigned ornithine (Orn) to CG and homoserine (Hsr) to AC, because these are intermediates in the pathways for synthesis of Arg and Thr, respectively, and are plausible early amino acids that might have later been supplanted by more complex amino acids. The assignments of Ser and Arg to AG and of Leu to CU likely reflect codon reassignments. Because Ser is synthesized from pyruvate, U rather than A should be the first base in its codons and, indeed, four of six codons for Ser in the modern code begin with U. Arg is synthesized from α-KG, and thus C rather than A should be the first base in its codons; four of six codons for Arg begin with C. A plausible primordial assignment for AG is 2,4-diaminobutyrate (Dab), which would be formed from OAA by reactions homologous to those that form ornithine from α-KG. The assignment of Leu to CU is also likely to reflect a codon reassignment. Because Leu is synthesized from pyruvate, U should be the first base in its codons. We have not postulated an early assignment for CU or for UA, which in the modern code corresponds to Tyr and the stop codons UAA and UAG. Tyr was unlikely to have been available in the earliest times, and a stop codon would have become relevant only after the emergence of translation.

Of the 16 entries in the postulated early doublet code, only two positions are unassigned. The remaining assignments respect a correlation between the first base of the codon and the α-keto acid precursor of the amino acid perfectly. Thus, the structural and synthetic homologies in this doublet code are stronger than those in the modern triplet code.

### A Model for Catalysis of Amino Acid Synthesis by Dinucleotides

We propose that catalysis of the reactions required to synthesize amino acids occurs in a complex in which an α-keto acid is esterified

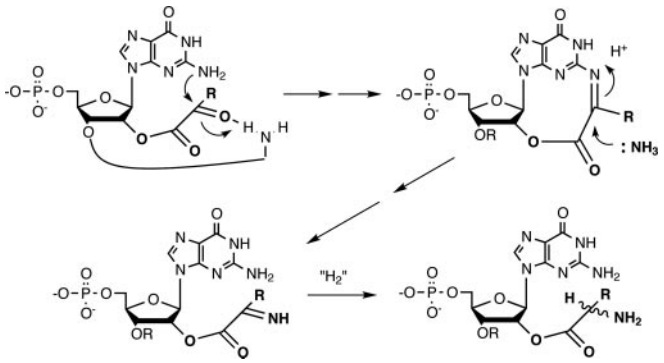

**Fig. 4.** Reductive amination of a covalently attached α-keto acid catalyzed by G in the first position of the dinucleotide catalyst. The curved line in the first structure and R in subsequent structures indicate the second nucleotide of the dinucleotide.

to the 2′ hydroxyl of the 5′ nucleotide of a dinucleotide (see Fig. 1). The feasibility of esterification at the internal 2′ hydroxyl is demonstrated by previous work showing that reaction of dinucleotides with activated amino acids gives acylation at the internal 2′ and terminal 2′(3′) positions in ratios ranging from 1:9 to 6:4, depending on the dinucleotide and the buffer concentration (38–40). The esterification reaction need not be specific, because reactions between an α-keto acid and a dinucleotide that cannot catalyze the first step in a synthetic pathway will simply lead to dead-end products. Thus, we do not need to invoke strong and specific binding interactions between α-keto acids and dinucleotides.

Only three types of reactions are required to initiate the conversion of α-keto acid precursors into the 15 amino acids shown in Fig. 3. These include (*i*) reductive amination of glyoxalate, pyruvate, OAA, or α-KG to give Gly, Ala, Asp, and Glu, respectively; (*ii*) phosphorylation of pyruvate, OAA, or α-KG in the pathways leading to Ser, Cys, homoserine, 2,4-diaminobutyrate, Asn, ornithine, Pro, and Gln; and (*iii*) decarboxylation of pyruvate and subsequent attack on either pyruvate or α-ketobutyrate in the pathways leading to Val, Leu, and Ile. In each case, the patterns in Fig. 3 and consideration of the mechanisms of the required reactions suggest that certain dinucleotides will be best able to catalyze the transformations that convert an α-keto acid into an amino acid, in the case of codons beginning with G, or into the first intermediate in a longer pathway, in the case of codons beginning with C, A, or U. The specific models for catalysis of reactions by certain dinucleotides that will be discussed below were developed by using molecular models to allow visualization of the conformations required to bring reacting atoms into proximity of potential catalytic groups.

### Pathways That Require Only Reductive Amination of an α-Keto Acid

Amino acids whose codons begin with G can be formed directly from simple α-keto acids by reductive amination. We propose that this reaction is catalyzed by the exocyclic amino group of G in the first position and the exocyclic amino groups of G, C, or A in the second position of the dinucleotide (see Fig. 4). The attack of the exocyclic amine of G on the carbonyl of an attached α-keto acid could be facilitated by a hydrogen bond between the carbonyl oxygen and an exocyclic amino group of the second base of the dinucleotide. Formation of the Schiff base, followed by attack of $NH_3$ and reduction of the resulting imine, would complete the synthesis of the amino acid. Notably, G is the only base that has an exocyclic amine in a position to reach the carbonyl of an α-keto acid at the 2′ position, providing a mechanistic interpretation for the association of G with these amino acids. Although the exocyclic amino groups of nucleobases are less nucleophilic than aliphatic amines due to resonance interactions with the aromatic ring, reactions of the exocyclic amine of G with ketones have been reported to result in both hydroxymethyl (41) and Schiff base (42) adducts. No further chemical reactions occur in the cases of Gly and Ala because these amino acids have no reactive positions. In the case of Asp and Glu, which could in theory undergo phosphorylation of the carboxylate (a reaction that occurs in many pathways originating from α-keto acids; see below), some factor must prevent further reaction. Possibly hydrogen bond interactions between the carboxylates of Asp and Glu and the exocyclic amines of the G and A sequester the carboxylate and disfavor the phosphoryl transfer reaction.

### Pathways That Begin with Phosphoryl Transfer

Pathways for synthesis of many amino acids begin with phosphorylation of α-KG, OAA, or pyruvate. We propose that phosphoryl transfer occurs from the 5′ position of the first nucleotide. Phosphoryl transfer should be facilitated in the covalent complex because it is intramolecular and because functional groups on the bases can orient the phosphoryl group and/or stabilize the transition state by polarizing the phosphate that is being attacked by the nucleophile or by stabilizing the developing negative charge on the leaving group as the P–O bond is cleaved. The catalytic abilities of the four bases will differ depending on the availability of hydrogen bond donors or acceptors within an appropriate distance and at an appropriate angle with respect to the transition state.‖

In the case of α-KG, a hydrogen bond between the exocyclic amine of C and an oxygen of the β phosphate should accelerate the reaction by orienting the phosphate and increasing the electrophilicity of the phosphorous (see Fig. 5*A*). Formation of a comparable hydrogen bond with the exocyclic amine of A is difficult, but an alternative orientation would allow a hydrogen bond between a hydroxyl of the β phosphate and N7 of A (see Fig. 5*B*). This interaction should have a lesser effect on the electrophilicity of the phosphorous. U has no exocyclic amine; orientation of the ring so that N3 can serve as a proton donor requires an unfavorable syn orientation of the base with respect to the ribose and furthermore results in an unfavorable interaction between the C2 carbonyl and the α phosphate (see Fig. 5*C*). Thus, the expected order of reactivity is C>A>U. If the second base of the dinucleotide is C, G, or A, interactions between the exocyclic amino group and the α phosphate could also assist the cleavage of the P–O bond. Thus, both bases may participate in the reaction, although specificity will be determined primarily by the first base.

A comparable analysis for OAA suggests that phosphoryl transfer will be most favorable when the first base of the dinucleotide is A. In this case, a hydroxyl on the β phosphate can be positioned to form a hydrogen bond with N7 of A (see Fig. 5*D*). Because the alkyl chain of OAA is shorter than that of α-KG, an interaction between the β phosphate and the exocyclic amine of C in the first position like that shown in Fig. 5*A* cannot be achieved. Interactions with U are unfavorable for the same reason given above. Thus, the expected order of reactivity is A>C, U. Again, exocyclic amino groups of C, G, or A in the second position of the dinucleotide could further facilitate phosphoryl transfer.

The first step in the pathway for synthesis of Ser and Cys** is phosphorylation of pyruvate to form phosphoenolpyruvate (PEP),

‖The structure of the transition state varies somewhat depending on whether a β or γ phosphate is being transferred. The following discussion pertains to transfer of a β phosphate.

**There are three pathways for Cys synthesis in extant organisms. Our analysis assumes that Cys is formed through Ser via attack of sulfide on *O*-acetylserine (as in *Escherichia coli*) or 3-phosphoserine, a plausible prebiotic intermediate.

EVOLUTION

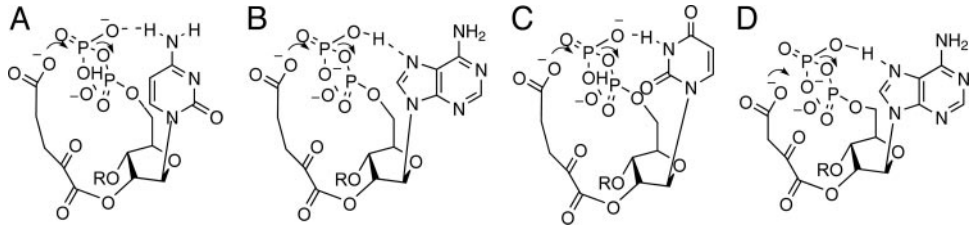


**Fig. 5.** Potential interactions between nucleobases and the transition state for phosphoryl transfer from the 5′$\beta$ phosphate to the carboxylate of $\alpha$-KG (*A–C*) or OAA (*D*) esterified to the internal 2′ hydroxyl of a dinucleotide. The nucleobase is C in *A*, A in *B*, U in *C*, and A in *D*. The second base is not shown for clarity.

a more difficult reaction because an enolate must be generated to provide a site for phosphorylation. Catalysis of this reaction is particularly important because PEP is an intermediate in the synthetic pathways for many sugars as well as for Ser and Cys. A plausible mechanism for this reaction is shown in Fig. 6. The reaction is proposed to occur in a step-wise fashion, with stabilization of an initially formed enolate by $Mg^{+2}$ associated with the diphosphate moiety of the dinucleotide, followed by attack of the enolate on the $\beta$ phosphate. Stabilization of an enolate by metal ions is a common strategy in the enolase superfamily of protein enzymes (43). The reaction could be further catalyzed by hydrogen bonding between either C or G in the second position and the $\beta$ phosphate that is to be transferred (not shown). A might also facilitate the reaction, but there is no amino acid assigned to the UA codon. The only hydrogen bond donor on U is located between two carbonyls, and approach to the $\beta$ phosphate would cause repulsive electrostatic interactions between the lone pairs of the carbonyls and the oxygens of the $\alpha$ and $\beta$ phosphates. Thus, we would predict that phosphorylation of pyruvate would not occur in the UU complex. Indeed, the UU dinucleotide is associated with a different synthetic pathway to be discussed below.

### Pathways That Begin with Decarboxylation of Pyruvate

The pathways for synthesis of the branched chain amino acids Val, Leu, and Ile begin with the most complex of the reactions considered here. This reaction involves decarboxylation of one molecule of pyruvate, followed by nucleophilic attack on a second molecule of pyruvate, in the case of Val and Leu, or $\alpha$-ketobutyrate, in the case of Ile. Notably, U is found in the second position of the codons for each of these amino acids, suggesting that its presence may facilitate this reaction. This would provide a mechanistic explanation for the previously noted association of hydrophobic amino acids with codons having U in the second position (1). In modern enzymes, this transformation is carried out by using thiamin pyrophosphate. We propose that U could serve as a primitive cofactor to accomplish the same chemistry. A possible mechanism for synthesis of 2-acetolactate from two molecules of pyruvate as it might be catalyzed by a UU dinucleotide is shown in Fig. 7, which is published as supporting information on the PNAS web site. A comparable mechanism can be proposed for the reaction involving $\alpha$-ketobutyrate attached to the 2′hydroxyl of the first nucleotide in an AU dinucleotide.

An interesting aspect of the doublet code in Fig. 3 is that Val and Leu are associated with different codons (GU and UU, respectively), yet both synthetic pathways begin with the same reactions, diverging only at the stage of 2-ketoisovalerate. We propose that both GU and UU can catalyze formation of 2-ketoisovalerate, but that the G in GU is able to catalyze reductive amination of 2-ketovalerate, leading to Val by the mechanism shown in Fig. 4. Because this reaction cannot be catalyzed by UU, additional reactions necessary to proceed further to the $\alpha$-keto acid precursor of Leu can occur before a final slow reductive amination to yield Leu.

### Subsequent Steps

The previous section describes plausible mechanisms for catalysis of the initial step in amino acid synthesis by various dinucleotides. For Gly, Ala, Glu, and Asp, only a single step is required. For the other amino acids, further elaboration of the side chain must occur. We postulate, as above, that the nature of the first and second bases of the dinucleotide will influence the rates of subsequent reactions, and that clues to the nature of favored reactions can be gained from correlations between the presence of certain bases and the occurrence of certain types of chemical reactions. We will discuss three examples to illustrate additional roles that could be played by dinucleotide catalysts.

For pathways originating from $\alpha$-KG and OAA, the presence of A in the second position of the codon is correlated with the formation of an amide from an acyl-phosphate precursor (see Fig. 2). (Note that an acyl phosphate intermediate is not formed in the pathways originating from pyruvate, because it lacks a terminal carboxylate, and there is no assigned amino acid for the UA position.) A possible explanation is that A in the second position catalyzes attack of $NH_3$ on the acyl-phosphate intermediate by removing a proton from $NH_3$ as it attacks the carbonyl of the acyl phosphate and by polarizing the carbonyl via a hydrogen bonding interaction with the exocyclic amine. The $pK_a$ of N1 of A in AMP is 3.9 (44), so the $pK_a$ would have to be perturbed for general base catalysis to occur at neutral pH. Significant perturbation of the $pK_a$s of A residues in RNAs has been reported (45, 46). C also has an exocyclic amine adjacent to an amide nitrogen that could serve as a general base catalyst, so C might also catalyze this reaction. The relative abilities of A and C to catalyze this reaction could be influenced by the greater size of A, the ability of A to adopt both syn and anti conformations with respect to the ribose ring while C must adopt an anti conformation, and the presence of the carbonyl adjacent to the amide nitrogen in C that is postulated to act as the general base. Experimental investigations will be required to determine whether A or C is a better catalyst for this reaction.

A particularly interesting transformation is the isomerization of 2-acetolactate or 2-aceto-2-hydroxybutyrate required for the synthesis of the branched chain amino acids. In extant enzymes, this isomerization is accomplished by two active-site $Mg^{2+}$ ions (47). One serves as a Lewis acid to polarize the carbonyl, whereas the second coordinates a hydroxyl group and may provide the hydroxide that deprotonates the substrate hydroxyl and initiates the isomerization reaction. An analogous reaction might be catalyzed by a $Mg^{2+}$ ion or ions coordinated to phosphate groups of a dinucleotide (see Fig. 8, which is published as supporting information on the PNAS web site).

The reductive amination that converts a completed $\alpha$-keto acid into an amino acid is particularly important because it establishes the stereochemical configuration (i.e., L or D) of the amino acid. A mechanism for reductive amination involving G in the first position of the dinucleotide was described above (see Fig. 4). For amino acids encoded by codons beginning with A, C, or U, this reaction must occur after some or all of the side-chain elaboration has taken place. In particular, it is necessary to avoid generation of the amino group while there are reactive groups present in the side chain, to

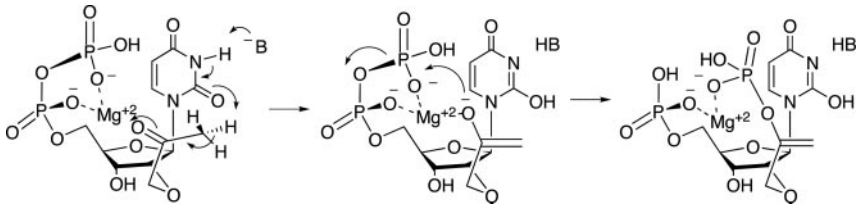

**Fig. 6.** Potential mechanism for formation of phosphoenolpyruvate catalyzed by UpU.

avoid the cyclization that occurs during synthesis of Pro (see Fig. 2). The final reductive amination could be catalyzed by A, G, or C in the second position, although probably not as effectively as by G in the first position. U does not have an exocyclic amine, and whether U can catalyze this reaction using the secondary amine in the ring is uncertain. Because the reaction is postulated to occur in a chiral environment, it would be expected to produce an excess of one stereoisomer. Thus, this model provides a plausible and testable hypothesis for the dominance of L-amino acids, a problem that has challenged prebiotic chemists for decades.

### The Possibility of Multiple Turnovers

The dinucleotide "catalysts" described above could be used for multiple turnovers if completed amino acids were removed by hydrolysis or transfer to some type of acceptor. Esterification with a new $\alpha$-keto acid could then initiate a new set of synthetic reactions. Because of the specificity proposed to reside in the complementarity between the particular $\alpha$-keto acid and the positions of the functional groups needed to catalyze the first reaction in a particular pathway, reaction with the "wrong" $\alpha$-keto acid would be unproductive and would tie up the catalyst in a nonproductive complex, at least until the $\alpha$-keto acid was removed by hydrolysis.

### Steps Toward Translation

The model presented here postulates that an association between amino acids and the first two bases of their codons arose before the emergence of macromolecules. The emergence of translation was obviously associated with expansion to a triplet code and selective pressures that led to codon assignments using the third position that minimize susceptibility to adverse effects of mutation and errors in translation. Furthermore, translation requires an association of amino acids with their anticodons, not with their codons. There are many ways in which these next steps toward translation might have occurred, and we have not yet examined these possibilities in detail. One intriguing possibility is that amino acids might be removed from their dinucleotide catalysts by transesterification to the 2′ hydroxyl of an RNA oligonucleotide. If this oligonucleotide were to recognize the base-pairing surface of the dinucleotide with a complementary sequence, then transesterification would lead to attachment of an amino acid to an RNA containing its anticodon. This would result in an early version of a charged "tRNA." Furthermore, the base following the doublet anticodon would be equivalent to the third position of an anticodon in a triplet code in which there was as yet no information content associated with the third position.

This work was supported by core funding from the Santa Fe Institute, a cooperative agreement between the National Aeronautics and Space Administration Astrobiology Institute and the University of Colorado (to S.D.C.), Insight Venture Partners (to D.E.S.), and the John Templeton Foundation (to H.J.M.).

1. Taylor, F. J. R. & Coates, D. (1989) *BioSystems* **22,** 177–187.
2. Crick, F. H. (1966) *J. Mol. Biol.* **19,** 548–555.
3. Knight, R. D., Freeland, S. J. & Landweber, L. F. (1999) *Trends Biochem. Sci.* **24,** 241–247.
4. Sonneborn, T. M., Bryson, V. & Vogel, H. J., eds. (1965) in *Evolving Genes and Proteins* (Academic, New York).
5. Woese, C. R. (1967) in *The Genetic Code: The Molecular Basis for Genetic Expression* (Harper & Row, San Francisco).
6. Wong, J. T. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 1909–1912.
7. Woese, C. R., Dugre, D. H., Saxinger, W. C. & Dugre, S. A. (1966) *Proc. Natl. Acad. Sci. USA* **55,** 966–974.
8. Pelc, S. R. & Welton, M. G. (1966) *Nature* **209,** 868–872.
9. Dunnill, P. (1966) *Nature* **210,** 1267–1268.
10. Yarus, M., Caporaso, J. G. & Knight, R. (2005) *Annu. Rev. Biochem.*, in press.
11. Crick, F. H. (1968) *J. Mol. Biol.* **38,** 367–379.
12. Smith, E. & Morowitz, H. J. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 13168–13173.
13. Lengeler, J., Drews, G. & Schlegel, H. (1999) in *Biology of the Prokaryotes* (Blackwell, Oxford), p. 165.
14. Karp, P. D., Riley, M., Paley, S. M. & Pellegrini-Toole, A. (2002) *Nucleic Acids Res.* **30,** 59–61.
15. Jadhav, V. R. & Yarus, M. (2002) *Biochemistry* **41,** 723–729.
16. Kumar, R. K. & Yarus, M. (2001) *Biochemistry* **40,** 6998–7004.
17. Huang, F., Bugg, C. W. & Yarus, M. (2000) *Biochemistry* **39,** 15548–15555.
18. Zhang, B. & Cech, T. R. (1997) *Nature* **390,** 96–100.
19. McGinness, K. E. & Joyce, G. F. (2003) *Chem. Biol.* **10,** 5–14.
20. Jenne, A. & Famulok, M. (1998) *Chem. Biol.* **5,** 23–34.
21. Seelig, B. & Jäschke, A. (1999) *Chem. Biol.* **6,** 167–176.
22. Weber, A. L. (2001) *Origins Life Evol. Biosphere* **31,** 71–86.
23. Pizzarello, S. & Weber, A. L. (2004) *Science* **303,** 1151.
24. Takagi, Y., Ikeda, Y. & Taira, K. (2004) *Top. Curr. Chem.* **232,** 213–251.
25. Jones, F. D. & Strobel, S. A. (2003) *Biochemistry* **42,** 4265–4276.
26. Cassano, A. G., Anderson, V. E. & Harris, M. E. (2004) *Biochemistry* **43,** 10547–10559.
27. Nakano, S., Chadalavada, D. M. & Bevilacqua, P. C. (2000) *Science* **287,** 1493–1497.
28. Higuchi, T., Eberson, L. & Herd, A. K. (1966) *J. Am. Chem. Soc.* **88,** 3805–3808.
29. Weinger, J. S., Parnell, K. M., Dorner, S., Green, R. & Strobel, S. A. (2004) *Nat. Struct. Mol. Biol.* **11,** 1101–1106.
30. Trifonov, E. N. (2002) *Biophysics* **47,** 539–544.
31. Hartman, H. (1975) *Origins Life* **6,** 423–427.
32. Hartman, H. (1995) *J. Mol. Evol.* **40,** 541–544.
33. Davis, B. K. (1999) *Prog. Biophys. Mol. Biol.* **72,** 157–243.
34. Tohá, J. C., Donoso, R., Estay, M. & Diaz-Valdes, J. (1989) *Med. Hypotheses* **30,** 265–269.
35. Jimenez-Sanchez, A. (1995) *J. Mol. Evol.* **41,** 712–716.
36. Schultz, D. W. & Yarus, M. (1996) *J. Mol. Evol.* **42,** 597–601.
37. Osawa, S. & Jukes, T. H. (1989) *J. Mol. Evol.* **28,** 271–278.
38. Tsilevich, T. L., Tarusova, N. B. & Gottikh, B. P. (1975) *Izv. Akad. Nauk. SSSR, Ser. Khim.* **4**, 916–921.
39. Tarusova, N. B., Tilevich, T. L. & Gottikh, B. P. (1975) *Izv. Akad. Nauk. SSSR, Ser. Khim.* **1**, 135–138.
40. Profy, A. T. & Usher, D. A. (1984) *J. Mol. Evol.* **20,** 147–156.
41. McGhee, J. D. & von Hippel, P. H. (1975) *Biochemistry* **14,** 1281–1296.
42. Papoulis, A., Al-Abed, Y. & Bucala, R. (1995) *Biochemistry* **34,** 648–655.
43. Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer, D. R., Barrett, W. C., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L. & Gerlt, J. A. (1996) *Biochemistry* **35,** 16489–16501.
44. Saenger, W. (1984) in *Principles of Nucleic Acid Structure*, ed. Cantor, C. R. (Springer, New York).
45. Yarus, M. Y. & Connell, G. J. (1994) *Science* **264,** 1137–1141.
46. Legault, P. & Pardi, A. (1994) *J. Am. Chem. Soc.* **116,** 8390–8391.
47. Dumas, R., Biou, V., Halgand, F., Douce, R. & Duggleby, R. G. (2001) *Acc. Chem. Res.* **34,** 399–408.