Norvaline and Norleucine May Have Been More Abundant Protein Components during Early Stages of Cell Evolution

3 authors:

Claudia Alvarez-Carreño
University College London
14 PUBLICATIONS   140 CITATIONS

Arturo Becerra
Universidad Nacional Autónoma de México
87 PUBLICATIONS   1,273 CITATIONS

Antonio Lazcano
Universidad Nacional Autónoma de México
253 PUBLICATIONS   7,509 CITATIONS

PROTEIN EVOLUTION

Norvaline and Norleucine May Have Been More Abundant Protein Components during Early Stages of Cell Evolution

Claudia Alvarez-Carreño · Arturo Becerra ·
Antonio Lazcano

Abstract The absence of the hydrophobic norvaline and norleucine in the inventory of protein amino acids is readdressed. The well-documented intracellular accumulation of these two amino acids results from the low-substrate specificity of the branched-chain amino acid biosynthetic enzymes that act over a number of related α-ketoacids. The lack of absolute substrate specificity of leucyl-tRNA synthase leads to a mischarged norvalyl-tRNA[Leu] that evades the translational proofreading activites and produces norvaline-containing proteins, (cf. Apostol et al. J Biol Chem 272:28980–28988, 1997). A similar situation explains the presence of minute but detectable amounts of norleucine in place of methionine. Since with few exceptions both leucine and methionine are rarely found in the catalytic sites of most enzymes, their substitution by norvaline and norleucine, respectively, would have not been strongly hindered in small structurally simple catalytic polypeptides during the early stages of biological evolution. The report that down-shifts of free oxygen lead to high levels of intracellular accumulation of pyruvate and the subsequent biosynthesis of norvaline (Soini et al. Microb Cell Factories 7:30, 2008) demonstrates the biochemical and metabolic consequences of the development of a highly oxidizing environment. The results discussed here also suggest that a broader definition of biomarkers in the search for extraterrestrial life may be required.

Keywords Norvaline · Norleucine · Prebiotic amino acids · Misincorporation of norvaline and norleucine in proteins · Aminoacyl-tRNA synthases

Introduction

A major unsolved issue in the study of the origin of life is the nature of the evolutionary processes that led to the selection of the L α-amino acids found in proteins from the large pool of prebiotic compounds. The ease of their formation in one-pot reactions such as those in Miller-Urey type experiments suggest they were present in the primitive Earth, a possibility strongly supported by the
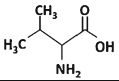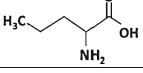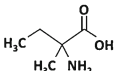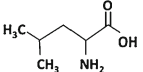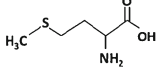
C. Alvarez-Carreño · A. Becerra · A. Lazcano (✉)
Facultad de Ciencias, Universidad Nacional Autonoma de Mexico, Apdo. Postal 70-407,
Cd. Universitaria, 04510 Mexico D.F, Mexico
e-mail: alar@ciencias.unam.mx

chemical characterization of approximately 80 different amino acids in carbonaceous chondrites (Burton et al. 2012). However, the fact that a number of amino acids found in contemporary life forms can be synthesized non-enzymatically under laboratory conditions or are present in meteorites does not necessarily imply that they were also essential for the appearance of life.

It is also possible that not all amino acids present in proteins today were available in the prebiotic environment but may be in fact the outcome of biological evolution. For instance, although a non-enzymatic synthesis of histidine has been reported (Shen et al. 1990), His may be a vestigial remnant of a catalytic nucleotide from the RNA World (White III 1976). Arginine and lysine, which are essential in nucleic acid-binding domains, are conspicuously absent both in carbonaceous chondrites and in prebiotic experiments (McDonald and Storrie-Lombardi 2010). Some of these compounds may have formed in minute amounts in the parental bodies of meteorites but have since decomposed (cf. Cleaves 2010). This possibility finds some support in the presence of small amounts of D/L ornithine, which is a product of arginine decomposition, in the preserved sample extracts from the Miller's simulation of the action of spark discharges over the reduced mixture of volcanic gases during prebiotic times (Johnson et al. 2008).

In an insightful paper published over 30 years ago, Weber and Miller (1981) discussed the occurrence of the twenty α-amino acids found in proteins. As they underlined, it is quite surprising that α-aminobutyric acid, α-aminoisobutyric acid, norvaline, alloisoleucine, norleucine, and homoserine, some of which are major prebiotic products, are missing in extant proteins (Table 1). Specifically, they raised the issue of the absence of hydrophobic amino acids including norvaline (Nva, 2-aminopentanoic acid) and norleucine (Nle, 2-aminohexanoic acid), which as they noted, "is most striking and a major challenge to any attempt to account for the selection of the twenty protein amino acids" (Weber and Miller 1981, p. 227). Weber and Miller (1981) suggested that this could be due to norvaline's enhanced rotational freedom around the protein backbone compared to that of valine. As noted by Weber and Miller, this could alter the conformation of proteins or prevent their proper folding, but concluded that it was not a

Table 1 Prebiotic synthesis of methionine and some protein and non-protein hydrophobic amino acids*

Name	Amino acid structure	Yields in spark discharge experiments (H₂O, CH₄, N₂, traces of NH₃) (Moles, relative to glycine=1) Ring et al 1972	Yields in volcanic spark discharge experiments (H₂O, CH₄, H₂O, NH₃) (Moles, relative to glycine=1) Johnson et al 2008*	Yields in spark discharge experiments with H₂S (H₂O, H₂S, CH₄, CO₂, NH₃) (Moles, relative to glycine=1) Parker et al 2011a*
Valine (Val) L-2-aminovaleric acid; 2-amino-3-methylbutanoic acid		10^{-2}	10^{-4}	0.8
Norvaline (Nva) L-2-aminovaleric acid; 2-aminopentanoic acid		10^{-1}	10^{-5}	not determined
Isovaline (Iva) 2-amino-2-methylbutanoic acid; 2-ethylalanine		10^{-2}	10^{-3}	10^{-2}
Leucine (Leu) L-2-amino-4-methylvaleric acid; 2-amino-4-methylpentatoic acid		10^{-2}	not determined	10^{-2}
Methionine (Met) L-2-amino-4-(methylthio)butyric acid; 2-amino-4-(methylsulfanyl)butanoic acid		--	--	10^{-3}

*Since the analytical procedure involved freeze-drying, the original volumes and the absolute relative yields are unknown

compelling explanation and suggested that perhaps the absence of norvaline was the outcome of chance events or a frozen coding accident.

As discussed below, a number of recent developments allow a reexamination of this issue based on the evidence of the presence of norvaline and norleucine in extant proteins. This may provide some insights that may change our understanding of the composition of primitive proteins. We discuss separately the case of norvaline and norleucine.

L-Norvaline is Incorporated into Proteins

Detectable amounts of L-norvaline were first reported in a serine-rich antifungal peptide produced by *Bacillus subtilis* which also included Asp, Glu, Thr, Ala, Tyr and Phe (Nandi and Sen 1953), and in *Serratia marcescens* regulatory mutants of leucine biosynthesis (Kisumi et al. 1976). These observations were confirmed by reports of the presence of small amounts of norvaline instead of leucine in the α and β subunits of recombinant human hemoglobin produced by *Escherichia coli* (Apostol et al. 1997).

It is now well established that together with norleucine and homoisoleucine, the intracellular accumulation of norvaline results from the low-substrate specificity of the branched-chain amino acid pathway enzymes (Kisumi et al. 1976; Bogosian et al. 1989). As is typical of a number of enzymes, substrate specificity is far from absolute, and many of the branched-chain amino acid biosynthetic enzymes can act over a number of related α-ketoacids. There is conclusive experimental evidence demonstrating that in recombinant strains of *E. coli* both norvaline and norleucine are synthesized by chain elongation from pyruvate-derived α-ketobutyrate as an alternate substrate of α-isopropylmalate synthase (EC 2.3.3.13). As demonstrated by Apostol et al. (1997), pyruvate is the primary substrate in the biosynthesis of both Nva and Nle. High concentrations of intracellular pyruvate lead to the rapid synthesis and accumulation of norvaline, which at first is similar to that reported for norleucine, but after 5 h Nva accumulates 3 or 4 times faster than norleucine (Apostol et al. 1997).

Recent experimental evidence has shown that the in vivo production and accumulation of norvaline is rapidly enhanced under low oxygen pressures in a non-recombinant *E. coli* strain (Soini et al. 2008). Under anaerobic conditions, such as those that likely existed in the primitive environment prior to the development of an oxidizing atmosphere, high glucose concentrations lead to a rapid accumulation of pyruvate, which is immediately used as an alternative substrate for direct keto chain elongation to α-ketobutyrate first, and then to α-ketovalerate which undergoes transamination and forms L-norvaline (Soini et al. 2008).

Measurements of amino acid concentrations in clarified crude broth lysates containing medium and cells indicated after 10 h a norvaline concentration of 1.0 mM, as high as that observed for aspartic acid (Soini et al. 2008), i.e., the enzyme-mediated production of norvaline is as efficient as that of Asp, a proteinogenic amino acid. The intracellular accumulation of pyruvate under the anaerobic conditions and high levels of glucose described by Soini et al. (2008) or due to other universally distributed pathways such as alanine transamination may be interpreted as an ancient process that may have taken place during the early stages of metabolic evolution.

The misincorporation of norvaline for leucine in recombinant hemoglobin is correlated with the ratio of free norvaline/leucine, suggesting that its presence is due to the misaminoacylation of tRNA[Leu] (Apostol et al. 1997). Although both leucine and norvaline are good helix formers (Lyu et al. 1991; Padmanabhan and Baldwin 1994), the incorporation of Nva in place of Leu in recombinant hemoglobin is not randomly distributed. This suggests that norvaline in certain positions of the recombinant hemoglobin may affect structural requirements for protein folding, insertion of the heme group or the assembly of the functional enzyme (Apostol et al. 1997).

The complex structure of extant enzymes is the evolutionary outcome of a complex series of events including internal gene duplications, slippage, accretion of functional domains and fusion events. It is likely that the earliest genetically encoded enzymes were small polypeptides with relatively simple structure, and that the basic traits of primitive active sites have been preserved during evolution and are among the oldest components of universally distributed extant enzymes. Although together with other nonpolar amino acids leucine plays a key role in aggregation processes that lead to proper protein folding, with few exceptions, it is conspicuously absent in the catalytic sites of most enzymes (Holliday et al. 2009). As shown in Table 2, only 15 of the 335 enzyme crystal structures available in the MACiE database (http://www.ebi.ac.uk/thornton-srv/databases/MACiE/queryMACiE.html), which includes detailed information of the catalytic activity of every Enzyme Commission (EC) enzyme subclass, have Leu in their catalytic site. Two of these 15 enzymes (malonyl-CoA-acyl carrier protein transacylase and phosphoenolpyruvate mutase) may be present in the three major cellular lineages due to horizontal gene transfer (HGT) events, and others have a restricted taxonomic distribution or are oxygen-dependent, which suggests that they are relatively recent developments (Table 2). As discussed below, this data suggests that leucine's catalytic activities are rather limited, and that its substitution by the less hydrophophic norvaline in small structurally simple primitive polypeptides during early biochemical evolution would have affected mostly their folding properties.

Branched-Chain Amino Acid Aminoacyl-tRNA Synthases Err

The fidelity of protein synthesis is dependent on accurate substrate recognition by the aminoacyl-tRNA synthetase (aaRS), which catalyzes the first step reaction of protein biosynthesis. Each aaRS recognizes a single cognate amino acid and covalently attaches it to the correct tRNA. The key role of the editing activity in recognizing the amino acid that corresponds to the precise anticodon is a striking demonstration of the dependency of the genetic code on the aaRS proofreading and editing of their mistakes (Döring et al. 2001; Ribas de Pouplana and Schimmel 2004).

However, such precision is far from absolute and in fact varies for different amino acids among the different aaRS (Young et al. 2011). Many aaRS cannot discriminate between cognate amino acids or structurally similar non-cognate amino acids in the synthetic reaction (Budisa 2004; Yadavalli and Ibba 2012; Cvetesic et al. 2013). This is the case of the branched aliphatic amino acids. Leucyl-, isoleucyl-and valyl-tRNA synthases (LeuRS, IleRS and ValRS, respectively) are members of an ancient homologous set of enzymes that are part of the class I subgroup of aaRS endowed with editing activities that hydrolyze mischarged tRNAs (Brown and Doolittle 1995). Their homologous editing domains share the same fold that must have appeared soon after the aaRS catalytic core structure was established, but prior to the duplications and divergence events that led to the three synthases (Cusack et al. 2000; Ribas de Pouplana and Schimmel 2004; Zhu et al. 2007). This evolutionary sequence of events can be understood in terms of the selection pressure imposed by the loose chemical specificity of primitive enzymes.

As Apostol et al. (1997) concluded, the incorporation of norvaline into proteins results from the misaminoacylation of tRNA[Leu]. Recent experimental data have demonstrated that the *E. coli* elongation factor Tu cannot discriminate against mischarged tRNA[Leu], as it binds with similar affinities to both leucyl-tRNA[Leu] and the mischarged valyl-tRNA[Leu], putting the burden of the editing activity on the aaRS itself (Cvetesic et al. 2013). Since the norvaline thee-carbon side chain cannot be excluded from the mischarged LeuRS, evasion of the translational proofreading activities leads to norvaline-containing proteins. As noted above,

Table 2 Enzymes with leucine residues in their reactive center (based on Holliday et al. ([2007](), [2009]()))

Enzyme name and EC number[a]	Reaction[b]	PDB ID	Amino acid residue[a]	Location of function[a]	Activity[a, d]	Metabolism[c]	Distribution[c]
camphor 5-monooxygenase EC 1.14.15.1	putidaredoxin + (+)-camphor + oxygen = (+)-exo-5-hydroxycamphor + oxidized putidaredoxin + H_2O	1yrc	Leu358	main chain amide	spectator	—	B
succinate dehydrogenase (ubiquinone) EC 1.3.5.1	ubiquinone + succinate = ubiquinol + fumarate	1yq4	Leu263, chainA	side chain	spectator	citrate cycle (TCA cycle); oxidative phosphorylation; biosynthesis of secondary metabolites	B, A, E
malonyl-CoA-acyl carrierprotein transacylase EC 2.3.1.39	malonyl-CoA + acyl-carrier protein = CoA + malonyl-[acyl-carrier protein]	1mla	Leu93	main chain amide	spectator	fatty acid biosynthesis	B, E A (HGT?)
histone acetyltransferase EC 2.3.1.48	acetyl-CoA + histone = CoA + acetylhistone	2gtr	Leu115, chainA Leu66, chainA	main chain amide main chain amide	spectator spectator	—	B, A, E
arallkylamine N-acetyltransferase EC 2.3.1.87	acetyl-CoA + 2-arylethylamine = CoA + N-acetyl-2-arylethylamine	1b6b	Leu111 Leu124	main chain carbonyl main chain amide	spectator spectator	tryptophan metabolism	E
D-alanine transaminase EC 2.6.1.21	D-alanine + 2-oxoglutarate = pyruvate + D-glutamate	1daa	Leu201	side chain	spectator	lysine degradation; arginine and proline metabolism; phenylalanine metabolism; D-arginine and D-ornithine metabolism; D-alanine metabolism	B
triacylglycerol lipase EC 3.1.1.3	triacylglycerol + H_2O = diacylglycerol + carboxylate	1hpl	Leu153	main chain amide	spectator	glycerolipid metabolism	B, E
peptide deformylase EC 3.5.1.88	formyl-L-methionyl peptide + H_2O = methionyl peptide + formate	1bsz	Leu91	main chain amide	spectator	—	B, E
tartronate-semialdehyde synthase EC 4.1.1.47	2 glyoxylate = 2-hydroxy-3-oxopropanoate + CO_2	2pan	Leu421, chainA Leu476, chain A	side chain side chain	spectator spectator	glyoxylate and dicarboxylate metabolism	B
anthranilate synthase EC 4.1.3.27	chorismate + L-glutamine = anthranilate + pyruvate + L-glutamate	1qdl	Leu86, chainB	main chain amide	spectator	phenylalanine, tyrosine and tryptophan biosynthesis; biosynthesis of secondary metabolites	B, A, E
trichodiene synthase EC 4.2.3.6	trans,trans-farnesyl diphosphate = trichodiene + diphosphate	1jfg	Leu97	side chain	spectator	sesquiterpenoid and triterpenoid biosynthesis	E

Table 2 (continued)

Enzyme name and EC number [a]	Reaction [b]	PDB ID	Amino acid residue [a]	Location of function [a]	Activity [a d]	Metabolism [c]	Distribution [c]
peptidylprolyl isomerase EC 5.2.1.8	peptidylproline (omega=180) = peptidylproline (omega=0)	1m9c	Leu122	side chain	spectator		B, A (?), E (?)
dodecenoyl-CoA isomerase EC 5.3.3.8	3-cis-dodecenoyl-CoA = 2-trans-dodecenoyl-CoA	1sg4	Leu66, chainA	main chain amide	spectator	fatty acid metabolism	B, E
phosphoenolpyruvate mutase EC 5.4.2.9	phosphoenolpyruvate = 3-phosphonopyruvate	1pym	Leu48	main chain amide	spectator	phosphonate and phosphinate metabolism; biosynthesis of secondary metabolites; microbial metabolism in diverse environments	B, E A (HGT?)
asparagine synthase (glutamine-hydrolysing) EC 6.3.5.4	ATP + L-aspartate + L-glutamine + H2O = AMP + diphosphate + L-asparagine + L-glutamate (overall reaction)	1ct9	Leu50	main chain carbonyl	spectator	alanine, aspartate and glutamate metabolism; nitrogen metabolism; biosynthesis of secondary metabolites; microbial metabolism in diverse environments	B, A, E

B bacteria; *A* archaea; *E* eucarya; — no entries; *HGT* horizontal gene transfer?

[a] (EMBL-EBI 2013); [b] McDonald (2013); [c] Kanehisa Laboratories (2012)

[d] A reactant is defined as an amino acid that undergoes a change in either electronic state or covalent bonding in any step of the reaction sequence in which it participates, and a spectator as an amino acid that exerts an electrostatic or steric effect upon another chemical species, which is important for the reaction to occur. This species may be another residue, water molecule, metal ion, substrate, intermediate or cofactor (Holliday et al. 2007, 2009; http://www.ebi.ac.uk/thornton-srv/databases/MACiE/glossary.html)

Table 3 Enzymes with methionine residues in their reactive center (based on Holliday et al. (2007, 2009))

Enzyme name and EC number [a]	Reaction [b]	PDB ID	Amino acid residue	Location of function	Activity [d]	Metabolism [c]	Distribution [c]
isovaleryl-CoA dehydrogenase EC 1.3.8.4	3-methylbutanoyl-CoA + electron-transferring flavoprotein = 3-methylcrotonyl-CoA + reduced electron-transferring flavoprotein	1ivh	Met135	main chain amide	spectator	valine, leucine and isoleucine degradation	B, A, E
nitrate reductase EC 1.7.99.4	nitrite + acceptor + H_2O = nitrate + reduced acceptor	1ogy	Met153, chainA Met346, chainA	side chain side chain	spectator spectator	nitrogen metabolism; microbial metabolism in diverse environments	B, A
plastoquinol-plastocyanin reductase EC 1.10.9.1	plastoquinol + 2 oxidized plastocyanin = plastoquinone + 2 reduced plastocyanin	2b3i	Met90, chainA	side chain	spectator	photosynthesis	B, A, E
chloride peroxidase EC 1.11.1.10	hydrogen peroxide + alkane + hydrochloric acid = RCl + 2 H_2O	1a7u	Met99	main chain amide	spectator	—	E A (HGT?) B (HGT?)
mercury(II) reductase EC 1.16.1.1	$Hg + NADP^+ + H^+ =$ mercury(2+) + NADPH	1zk7	Met9, chainA	side chain	spectator	—	B, A, E
4-cresol dehydrogenase (hydroxylating) EC 1.17.99.1	4-cresol + acceptor + H_2O = 4-hydroxybenzaldehyde + reduced acceptor	1dii	Met650, chainC	side chain	reactant	toluene degradation; microbial metabolism in diverse environments	B
acetolactate synthase EC 2.2.1.6	2-acetolactate + CO_2 = 2 pyruvate	1n0h	Met582, chainB	side chain	spectator	valine, leucine and isoleucine biosynthesis; butanoate metabolism; C5-branched dibasic acid metabolism; pantothenate and CoA biosynthesis; biosynthesis of secondary metabolites	B, A, E
pyruvate, phosphate dikinase EC 2.7.9.1	ATP + pyruvate + orthophosphate = AMP + phosphoenolpyruvate + diphosphate	1kc7	Met103	main chain amide	spectator	pyruvate metabolism; carbon fixation in photosynthetic organisms; microbial metabolism in diverse environments	B, A, E

Table 3 (continued)

Enzyme name and EC number[a]	Reaction[b]	PDB ID	Amino acid residue	Location of function	Activity[d]	Metabolism[c]	Distribution[c]
phosphonoacetaldehyde hydrolase EC 3.11.1.1	phosphonoacetaldehyde + H_2O = Acetaldehyde + Orthophosphate	1rql	Met49	side chain	spectator	phosphonate and phosphinate metabolism; microbial metabolism in diverse environments	B, E (HGT?)
xylose isomerase EC 5.3.1.5	D-xylose = D-xylulose	1xld	Met83, chainA	side chain	spectator	pentose and glucuronate interconversions; fructose and mannose metabolism	B, E A (HGT?)

B bacteria; *A* archaea; *E* eucarya; — no entries; *HGT* horizontal gene transfer?

[a] (EMBL-EBI 2013); [b] McDonald (2013); [c] Kanehisa Laboratories (2012)

[d] definitions of spectator and reactant as in Table 2

the available data shows that the misincorporation of norvaline for leucine is correlated with the ratio of free norvaline/leucine, suggesting that significant amounts of Nva can be incorporated in the hydrophobic regions of proteins. This possibility is amenable to experimental analysis with model polypeptides.

Size is Not Enough: Norleucine as a Substitute for Methionine in Protein Chains

Norleucine is also a by-product of the leucine biosynthetic pathway enzymes starting from pyruvate or α-ketobutyrate in place of α-ketoisovalerate, and can be misincorporated in place of methionine in recombinant proteins (Bogosian et al. 1989). Methionine and norleucine have comparable sizes and slightly different electric charges, but Met is much more hydrophilic than Nle (Table 4). It is known that the in vitro aggregation properties of human prion protein are dramatically altered by changing methionine for the more hydrophobic norleucine (Wolschner et al. 2009), and there is an ample bibliography that it can replace Met in a number of bacterial proteins, in many cases without loss of catalytic activity (Munier and Cohen 1959; Cohen and Munier 1959; Cowie et al. 1959; Anfinsen and Corley 1969; Old and Jones 1975). Like leucine, methionine is a good helix forming amino acid (Pace and Scholtz 1998), and is rarely found in the catalytic site of enzymes (Holliday et al. 2009). As shown in Table 3, only ten enzymes of the 335 present in the MACiE database, have methionine in their catalytic site, and only in one of these ten enzymes (4-cresol dehydrogenase) it partakes in catalysis. Of these ten entries, two of them (chloride peroxidase and xylose isomerase) may be present in the three primary kingdoms due to horizontal gene transfer (HGT) events, and three others (nitrate reductase, 4-cresol dehydrogenase, and phosphoenolacetaldehyde hydrolase) have a restricted taxonomic distribution or are oxygen-dependent, which suggests that they are relatively recent developments.

Norleucine can be substituted for methionine in recombinant bovine protein (Bogosian et al. 1989). The presence of Nle in the dimer interface of the human enteric α-defensin HD5, a multifunctional small antimicrobial peptide, leads to an atypical parallel mode of dimerization but without affecting its antibacterial activity (Rajabi et al. 2012). As suggested by Barker and Burton (1979), the toxic effects of norleucine on microbial growth (Richmond 1962) are due to its inhibition of methylation reactions. Indeed, in addition to its role as a building block in proteins, methionine is the immediate precursor of S-adenosylmethionine (SAM), which is one of the major methyl-group donors in trans-methylation reactions in contemporary biochemistry. Accordingly, the suggestion that ribonucleotide-like coenzymes are remnants of an ancestral stage in which ribozymes played a more conspicuous role in metabolism (Orgel and Sulston 1971; White III 1976) would imply that methionine was incorporated into biological systems mostly because of its involvement in methyltransferase activities that evolved in a primordial RNA-dependent world (Parker et al. 2011b).

Table 4 Hydrophobicity (log *P*), size (van der Waals volume) and electrical charge (*p* I) of hydrophobic amino acids*

Amino acid	log *P*	van der Waals volume ($Å^3$)	p *I*
Leucine	−1.59	138.27	6.195
Norvaline	−2.01	120.98	6.27
Methionine	−2.41	139.49	5.77
Norleucine	−1.52	138.27	6.15

Conclusions

As summarized by Cleaves (2010), the presence of the 20 protein α-amino acids defies a simple, lineal explanation and should be seen as the outcome of a complex combination of their prebiotic accessibility, their metabolic availability, and the selection advantages derived from their different functional groups. Evidence of the incorporation of norvaline and, to a lesser extent, norleucine and other amino acids, suggests that the lack of absolute substrate specificity that characterizes not only the enzymes of branched-chain amino acid biosynthesis but also a number of aaRS implies that their chemical ambiguity has played a role in defining the expression of the genetic code.

The high values of amino acid half-lives as compared to those of ribonucleotides and RNA (cf. Cleaves 2012) support the possibility that when protein biosynthesis first evolved amino acids of prebiotic origin were still available. The inventory of prebiotic compounds may have not included key amino acids like histidine but, as reviewed here, the available evidence suggests that, in addition to valine, leucine and isoleucine, a number of related compounds such as norvaline, norleucine, α-amino-n-butyric acid and alloisoleucine, among others, must have also been present (Table 1). The alcoholic fermentation of α-amino-n-butyric acid (Kepner et al. 1954) suggests that some of these compounds may have ended up as sources of carbon and energy for primordial heterotrophs.

The evidence that leucyl-, valyl- and isoleucyl-tRNA synthases are the outcome of serial gene duplications implies that they descend from a largely unspecific enzyme (Zhu et al. 2007) whose substrate ambiguity we hypothesize may have led to the incorporation of higher amounts of prebiotic norvaline in early proteins. Evolutionary analysis of substrate promiscuity in enzymes mediating physiological and metabolic adaptations demonstrates that is in many cases it represents the starting point towards the evolution of new catalytic activities (Tawfik and Khersonsky 2010; Tawfik 2013). As summarized by Liu and Schultz (2010), there is evidence that under in vivo or in vitro conditions the lack of absolute specificity of the elongation factor Tu and the ribosome can lead to the incorporation of noncanonical amino acids and other compounds, including D amino acids and α- hydroxy acids. These observations support Woese's contention that primitive protein synthesis was an error-prone process producing small catalytic peptides with somewhat different sequences lacking the accuracy and specificity of extant enzymes (Woese 1965, 1987).

The exhaustion of the prebiotic budget of norvaline, norleucine and other non-proteinic amino acids did not stop their misincorporation into proteins. The mechanisms described here may have operated during the early stages of biochemical evolution, but continued afterwards when the development of the biosynthesis of branched-chain amino acids led to norvaline and norleucine as by-products. The incorporation in proteins of norvaline in place of leucine is an outcome of the combination of the substrate ambiguity and multifunctionality of both leucyl-tRNA synthase and the branched-chain amino acid biosynthetic enzymes. Such functional flexibilities may confer evolutionary advantages, especially under anaerobic conditions that favor the accumulation of pyruvate.

The lack of absolute substrate specificity is neither primitive nor a specific adaptation, but the outcome of the inability of a number of enzymes to distinguish between related substrates. Comparative quantifiable data on chemical and physical properties of coded and non-coded amino acids provide important insights on the processes that led to the establishment of the standard genetic code (Lu and Freeland 2008; Philip and Freeland (2011). As summarized by Freeland and his coworkers, hydrophobicity, and charge of amino acids play a key role in protein folding and stability (cf. Lu and Freeland 2008). As shown in Table 4, although the electric charges of leucine and norvaline are comparable, Nva is not only significantly smaller that Leu, but also less hydrophobic. Protein folding is driven by

the formation of a hydrophobic core, and it is reasonable to assume that as enzymes became increasingly complex, replacing leucine with the smaller and less hydrophobic norvaline would affect the packing of secondary structural elements required for the formation of the core. It would also affect protein ligand binding and protein-protein interactions. It can be hypothesized that the lower hydrophobicity of norvaline would probably also limit its ability to form stabilizing structures such as the Leu-rich repeats and the Leu-zippers that play a key role in a number of DNA- and RNA-binding proteins.

The evidence that norvaline biosynthesis is enhanced under anaerobic conditions (Soini et al. 2008) is a surprising demonstration of the indirect biochemical and metabolic consequences of the planetary transition to a highly oxidizing environment (Lazcano 2012). It also suggests that the broad LeuRS editing properties of *Aquifex aeolicus* (Zhu et al. 2007), a deeply branching hyperthermophilic microaerophilic bacteria (Deckert et al. 1998), may be understood in terms of their role in limiting the misincorporation of norvaline and norleucine that accumulate because of enhanced anaerobic pyruvate production. The results reviewed here also suggest that the search for hypothetical extraterrestrial life forms should consider the possibility that other biochemistries (Bada 2001; Pace 2001) may be based on a wider range of monomers.

References

Anfinsen C, Corley L (1969) An active variant of staphylococcal nuclease containing norleucine in place of methionine. J Biol Chem 244:5149–5152

Apostol I, Levine J, Lippincott J et al (1997) Incorporation of norvaline at leucine positions in recombinant human hemoglobin expressed in *Escherichia coli*. J Biol Chem 272:28980–28988

Bada JL (2001) State-of-the-art instruments for detecting extraterrestrial life. Proc Natl Acad Sci U S A 98:707–800

Barker DG, Bruton CJ (1979) The fate of norleucine as a replacement for methionine in protein synthesis. J Mol Biol 133:217–231

Bogosian G, Violand BN, Dorward-King EJ et al (1989) Biosynthesis and incorporation into protein of norleucine by *Escherichia coli*. J Biol Chem 264:531–539

Brown J, Doolittle W (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthase gene duplications. Proc Natl Acad Sci U S A 92:2441–2445

Budisa N (2004) Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire. Angew Chem Int Ed 43:6426–6463

Burton AS, Stern JC, Elsila JE et al (2012) Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites. Chem Soc Rev 41:5459–5472

Cleaves HJ (2010) The origin of the biologically coded amino acids. J Theor Biol 263:490–498

Cleaves HJ (2012) Prebiotic chemistry: what we know, what we don't. Evol Educ Outreach 5:342–360

Cohen GN, Munier R (1959) Effects of structural analogues of amino acids on growth and synthesis of proteins and enzymes in *Escherichia coli*. Biochim Biophys Acta 31:347–356

Cowie DB, Cohen GN, Bolton ET, De Robichon-Szulmajster H (1959) Amino acid analog incorporation into bacterial proteins. Biochim Biophys Acta 34:39–46

Cusack S, Yaremchuk A, Tukalo M (2000) The 2 Å crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adenylate analogue. EMBO J 19:2351–2361

Cvetesic N, Akmacic I, Gruic-Sovulj I (2013) Lack of discrimination against non-proteinogenic amino acid norvaline by elongation factor Tu from *Escherichia coli*. Croat Chem Acta 86:73–82

Deckert G, Warren PV, Gaasterland T et al (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. Nature 392:353–358

Döring V, Mootz HD, Nangle LA et al (2001) Enlarging the amino acid set of Escherichia coli by infiltration of the valine coding pathway. Science 292:501–504

EMBL-EBI (2013) Query MACiE Reactions by Amino Acid Residue. http://www.ebi.ac.uk/thornton-srv/databases/MACiE/queryMACiE.html. Accessed 21 June 2013

Holliday GL, Almonacid DE, Mitchell JBO, Thorton JM (2007) The chemistry of protein catalysis. J Mol Biol 372:1261–1277

Holliday GL, Mitchell JBO, Thornton JM (2009) Understanding the functional roles of amino acid residues in enzyme catalysis. J Mol Biol 390:560–577

Johnson AP, Cleaves HJ, Dworkin JP et al (2008) The Miller volcanic spark discharge experiment. Science 322:404

Kanehisa Laboratories (2012) KEGG: Kyoto encyclopedia of genes and genomes. http://www.genome.jp/kegg/. Accessed 21 June 2013

Kepner RE, Castor JGB, Webb AD (1954) Conversion of α-amino-n-butyric acid to n-propanol during alcoholic fermentation. Arch Biochem Biophys 51:88–93

Kisumi M, Sugiura M, Chibata I (1976) Biosynthesis of norvaline, norleucine, and homoisoleucine in *Serratia marcescens*. J Biochem 80:333–339

Lazcano A (2012) Planetary change and biochemical adaptation: molecular evolution of corrinoid and heme biosynthesis. Hematol Suppl 1:S7–S10

Liu CC, Schultz PG (2010) Adding new chemistries to the genetic code. Annu Rev Biochem 79:413–444

Lu Y, Freeland SJ (2008) A quantitative investigaction of the chemical space surrounding amino acid alphabet formation. J Theor Biol 250:349–361

Lyu PC, Sherman A, Kallenbach NR (1991) α-Helix stabilization by natural and unnatural amino acids with alkyl side-chains. Proc Natl Acad Sci U S A 88:5317–5320

McDonald A (2013) ExplorEnz—the enzyme database. In: International Union of Biochemistry and Molecular Biology. http://www.enzyme-database.org/. Accessed 26 June 2013

McDonald GD, Storrie-Lombardi MC (2010) Biochemical constraints in a protobiotic earth devoid of basic amino acids: the "BAA(−) world". Astrobiology 10:989–1000

Munier R, Cohen GN (1959) Incorporation of structural analogues of amino acids into bacterial proteins during their synthesis *in vivo*. Biochim Biophys Acta 31:378–391

Nandi P, Sen GP (1953) An antifungal substance from a strain of *B subtilis*. Nature 172:871–872

Old JM, Jones DS (1975) The recognition of methionine analogues by Escherichia coli methionyl-transfer ribonucleic acid synthetase. Biochem Soc Trans 3:659–660

Orgel LE, Sulston JE (1971) Polynucleotide replication and the origin of life. In: Kimball AP, Oró J (eds) Prebiotic and biochemical evolution. North-Holland, Amsterdam, pp 89–94

Pace NR (2001) The universal nature of biochemistry. Proc Natl Acad Sci U S A 98:805–808

Pace CN, Scholtz JM (1998) A helix propentisty scale based on experimental studies of peptides and proteins. Biophys J 75:422–427

Padmanabhan S, Baldwin RL (1994) Test for helix-stabilizing interactions between various nonpolar side chains in alanine-based peptides. Protein Sci 3:1992–1997

Parker ET, Cleaves HJ, Dworkin JP et al (2011a) Primordial synthesis of amines and amino acids in a 1958 H2S-rich spark discharge experiment. Proc Natl Acad Sci U S A 108:5526–5531

Parker ET, Cleaves HJ, Callahan MP et al (2011b) Prebiotic synthesis of methionine and other sulfur-containing organic compounds on the primitive Earth: a contemporary reassessment based on an unpublished 1958 Stanley Miller experiment. Orig Life Evol Biosph 41:201–212

Philip GK, Freeland SJ (2011) Did evolution select a nonrandom "alphabet" of amino acids? Astrobiology 11:235–240

Rajabi M, Ericksen B, Wu X, de Leeuw E, Zhao L, Pazgier M, Lu W (2012) Functional determinants of human enteric a-defensin HD5: crucial role for hydrophobicity at dimer interface. J Biol Chem 287:21615–21627

Ribas De Pouplana L, Schimmel P (2004) Aminoacyl-tRNA synthetases as clues to the establishment of the genetic code. In: Ribas De Pouplana L (ed) The genetic code and the origin of life. Springer US, Georgetown, pp 119–133

Richmond MH (1962) The effect of amino acid analogues on growth and protein synthesis in microorganisms. Bacteriol Rev 26:398–420

Ring D, Wolman Y, Friedmann N, Miller SL (1972) Prebiotic synthesis of hydrophobic and protein amino acids. Proc Natl Acad Sci USA 69:765–768

Shen C, Yang L, Miller SL, Oró J (1990) Prebiotic synthesis of histidine. J Mol Evol 31:167–174

Soini J, Falschlehner C, Liedert C et al (2008) Norvaline is accumulated after a down-shift of oxygen in *Escherichia coli* W3110. Microb Cell Factories 7:30

Tawfik DS (2013) Messy biology and the origins of evolutionary innovations. Nat Chem Biol 6:692–696

Tawfik DS, Khersonsky O (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. Annu Rev Biochem 79:471–505

Weber A, Miller S (1981) Reasons for the occurrence of the twenty coded protein amino acids. J Mol Evol 17:273–284

White HB III (1976) Coenzymes as fossils of an earlier metabolic state. J Mol Evol 7:101–104

Woese CR (1965) On the evolution of the genetic code. Proc Natl Acad Sci U S A 54:1546–1552

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271

Wolschner C, Giese A, Kretzschmar HA et al (2009) Design of anti- and pro-aggregation variants to assess the effects of methionine oxidation in human prion protein. Proc Natl Acad Sci U S A 106:7756–7761

Yadavalli SS, Ibba M (2012) Quality control in aminoacyl-tRNA synthesis its role in translational fidelity. Adv Protein Chem Struct Biol 86:1–4

Young DD, Young TS, Jahnz M, Ahmad I, Spraggon G, Schultz PG (2011) An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. Biochemistry 50:1894–1900

Zhu B, Zhao M-W, Eriani G, Wang E-D (2007) A present-day aminoacyl-tRNA synthetase with ancestral editing properties. RNA 13:15–21