J Mol Evol (1987) 25:15–19

Journal of Molecular Evolution

β-Turn-Driven Early Evolution: The Genetic Code and Biosynthetic Pathways

Jerzy Jurka and Temple F. Smith

Dana–Farber Cancer Institute, Harvard University, 44 Binney Street, Boston, Massachusetts 02115, USA

Summary. The physicochemical properties of β-turns suggest their biological importance prior to the formation of the genetic code. These properties include ones potentially affecting the preference for either L- or D-amino acids. The abundance of certain amino acids in β-turns is correlated with their assignment to a small, well-defined part of the genetic code and with their role as metabolic precursors for other amino acids. It is proposed that in the prebiotic environment, β-turns became objects of selection that influenced the evolution of the genetic code and biosynthetic pathways for amino acids.

Key words: β-Turns — Secondary structures — Chirality — Genetic code — Biosynthetic pathways — Evolution

Introduction

The conformational flexibility of contemporary enzymes is restrained by their elaborate secondary and tertiary structures. The shorter the polypeptide chain is, the more difficult its flexibility is to stabilize. This creates obvious difficulties for everybody who tries to analyze the origin and evolution of the earliest catalytically active protein structures. Such structures, if formed of short peptides, would have had to be stable enough to participate in chemical reactions. Many researchers interested in this problem have focused their efforts on β-sheets as suitable structural elements (Orgel 1972, 1975; Brack and Orgel 1975; Brack 1977; von Heijne et al. 1978; Marlborough 1980). Orgel (1977) also discussed the importance of β-turns stabilized by β-sheets as plausible sites of the early enzymatic activity. However, those amino acids that are good β-sheet formers are also good β-turn breakers and vice versa. Introducing both structures, and thereby both kinds of amino acids, into the primitive peptides required quite elaborate replication and translation processes to avoid frequent confusion between β-turn formers and β-turn breakers. It is worth considering that even without β-sheets, β-turns could be stabilized by cyclization of peptides, formation of complexes with metals, and so forth. As discussed below, β-turns themselves may have been the dominant early structures. Furthermore, we find it useful to distinguish an era during early prebiotic history when β-turns, as the first bioactive protein structures, were a driving force behind the evolution of prebiotic systems toward modern life.

Stability and Optical Activity of β-Turns

β-Turns are well-defined secondary structures in which four consecutive residues fold the polypeptide chain back upon itself by nearly 180° (see Lewis et al. 1971, 1973). After Venkatachalam (1968), we distinguish three basic forms of β-turns on the basis of their conformations. They are denoted as I, II, and III and their mirror images as I′, II′, and III′. Other forms of β-turns have also been defined (Lewis et al. 1973), but they appear to be much less common and have been less thoroughly investigated. β-Turns are known to be preferred structures in

Offprint requests to: J. Jurka

Table 1. Amino acid composition of β-turns

Type II(II'), positions 2 and 3[a]		Overall[b]	
Amino acid	Frequency	Amino acid	Frequency
Gly	40	Gly	194
Ala	10	Ser	155
Ser	10	Asp	118
Asn	9	Asn	106
Asp	8	Lys	103
Lys	5	Ala	85
Pro	5	Pro	79
Arg	4	Thr	79
Met	3	Leu	62
Phe	3	Tyr	62
Thr	3	Val	53
Cys	2	Glu	51
Val	2	Gln	46
Gln	1	Arg	40
Glu	1	His	36
His	1	Cys	33
Tyr	1	Ile	32
Ile	0	Phe	30
Leu	0	Trp	22
Trp	0	Met	13

[a] Amino acid composition at positions 2 and 3 of type II(II') β-turns from globular proteins (after Smith and Pease 1980)
[b] Overall amino acid composition of β-turns for 29 proteins as compiled by Chou and Fasman (1978)

small linear and circular peptides (for reviews, see Smith and Pease 1980; Toniolo 1980; Rose et al. 1985). They were shown to be stable in a variety of environments, including strong solvents (Rose et al. 1985). Analysis of amino acid conformations at positions 2 and 3 of β-turns led to a fundamental distinction between homochiral and heterochiral sequences (Venkatachalam 1968). The former were predicted to favor I(LL) and I'(DD) and the latter II(LD) and II'(DL) forms of β-turns. Venkatachalam (1968) also predicted that, in type II(II') β-turns, glycine can either play the role of a D-residue if the other position is occupied by an L-residue or vice versa. Indeed, Gly is predominant at position 2 in type II β-turns from globular proteins. Similarly, in type II' β-turns Gly is found at position 3 in 14 of 16 cases (Smith and Pease 1980). Since positions 2 in type II and 3 in type II' β-turns were predicted to favor D-residues, Gly has been called a "pseudo-D" residue (Smith and Pease 1980).

Experimental analysis of model peptides has revealed more β-turns of type II(II') than of any of the remaining types of turns (see Rose et al. 1985). This suggests that the homochiral turns are less stable and raises the possibility that among the randomly synthesized β-turns, heterochiral types II(II') predominate over the homochiral types of turns. This can be expressed schematically as L(2)D(3) + D(2)L(3) > L(2)L(3) + D(2)D(3).

Given that polymerization of short peptides occurred in the prebiotic environment, formation of β-turns in these peptides could be possible. These early turns were likely to be rich in Gly, perhaps even more likely than are current turns, since Gly appears to be the most abundant amino acid in simulated prebiotic syntheses of amino acids (Harada and Fox 1964; Weber and Miller 1981). In this situation Gly(2)D,L(3) and L,D(2)Gly(3) might have become the most abundant among type II(II') β-turns. Accommodation of optically inactive Gly in either position 2 or 3 creates local optical asymmetry in potentially symmetrical forms of β-turns. It is tempting to speculate that this phenomenon might have promoted enrichment of the optical purity of β-turns. Another potentially important fact in this context is that in globular proteins, type II β-turns containing L(2)Gly(3) are almost twice as abundant as type II' containing Gly(2)L(3) (Table 11 from Smith and Pease 1980). A priori, one would expect them to be in the ratio 1:1. A trivial explanation of this inequality would be that the sequence L(2)Gly(3) is more stable than Gly(2)L(3), and therefore is selected for. Indeed, X–Gly and Gly–X dipeptides, where X is either an L- or a D-residue, differ in stability in the presence of metals (Siegel and Martin 1982). Whether the data for dipeptides can be extrapolated to the entire β-turn in globular proteins remains to be shown. A less trivial explanation would be that L-residues in positions 1 and 4 are more compatible with type II than with type II' β-turns. If so, one could expect this to be an additional factor promoting selection of optically pure forms of β-turns. This question can be addressed experimentally. Physicochemical selection of optically enriched forms of β-turns could be the first step toward evolution of L-proteins. This does not address the question of why L- and not D-proteins were finally selected. That may have been a matter of a chance (Wald 1957).

Relationships to the Genetic Code and Metabolic Pathways

Physicochemical restrictions imposed on positions 2 and 3 in type II(II') β-turns are reflected in the frequent occurrence of Gly in these positions (Table 1). We assume that proportions of other amino acids are also influenced by the inherent physicochemical properties of β-turns. If we are right, then the amino acids that are predominant in contemporary β-turns might have also been of importance in the earliest β-turns. The analysis presented further on is based on this assumption.

In Table 1 we list the overall amino acid frequencies from the β-turns of 29 proteins, as com-

Table 2. Relative numbers of amino acids from β-turns assigned to the genetic code[a]

First base	Second base			
	Y		R	
Y	Phe 15.00 Leu ( )	Ser 25.83	Tyr 31.00 His 18.00	Cys 16.50 Trp 22.00
	Leu 10.33	Pro 19.75	Gln 23.00	Arg 6.67
R	Ile 10.67 Met 13.00	Thr 19.75	Asn 53.00 Lys 51.50 Asp 59.00	Ser ? Arg ?
	Val 13.25	Ala 21.25	Glu 25.50	Gly 48.50

[a] The numbers from the right column of Table 1 were divided by the numbers of codons for the respective amino acids and assigned to a simplified scheme of the genetic code. R, purine; Y, pyrimidine

piled by Chou and Fasman (1978). The top five amino acids—Gly, Ser, Asp, Asn, and Lys—make up almost half of the total number of amino acids in all types of β-turns. Furthermore, they are assigned to the same general type of codon. To illustrate this, we divided the numbers from Table 1 by the codon degeneracy and assigned the resulting amounts to a simplified scheme of the genetic code (Table 2). The most abundant amino acids are located in the bottom right corner of Table 2, containing RRY and RRR types of codons. Amino acids coded for by RRY codons are more abundant than those coded for by RRR codons. This is due to the relatively low content of Arg and Glu in β-turns. Column S1 of Table 3 summarizes the number of amino acids per codon assigned to each "generalized codon." In column S2 we show the same numbers but with Ser and Arg excluded because each of these amino acids is assigned to two different types of codons: RRN and YYN; and RRN and YRN, respectively.

According to the classification by Chou and Fasman (1978), none of the RRN amino acids is a β-sheet former. More exactly, they are either indifferent or β-breakers. This provides another basis on which to relate them as a class to their structural role in proteins. Outside the RRN group, only histidine and proline are β-sheet breakers. Proline of course plays a very dominant role in many extant protein β-turns; however, there may be some justification in considering amino acids like proline, methionine, and tryptophan as late additions, given their unique structures and/or single-codon assignments.

In the biosynthetic pathways of amino acids, Ala, Asp, Glu, Gly, and Ser are mutually interconvertible (Wong 1975). With the exception of Glu, these amino acids are among the most abundant in β-turns

Table 3. Summarized relative numbers by codon type[a]

Codon type	S1	S2
RRN	270.0	237.5
YRN	117.7	111.0
RYN	77.9	77.9
YYN	70.9	45.1

[a] S1 denotes total relative number of amino acids assigned to the corresponding codon type. S2 is the same number after subtraction of Ser and Arg. R, purine; Y, pyrimidine; N, any base

(see Table 1). All but Ala are assigned to the RRN type of codons. Three of them, Asp, Glu, and Ser, are involved in the following precursor–product relations with other amino acids: Asp → Asn, Asp → Lys, Asp → Thr, Glu → Arg, Glu → Gln, Glu → Pro, Ser → Cys, Ser → Trp. There are no such one-directional precursor–product relations from non-RRN to RRN amino acids. From this we conclude that the biosynthesis of those amino acids that are among the most abundant in β-turns and are precursors to other amino acids was implemented very early in evolution. A driving force behind this process could have been selection for β-turns. The beginnings of the major biosynthetic pathways coincide with the amino acids coded for by RRN-type codons. The RRN amino acids as a class are also the most abundant in β-turns. Therefore, we find the RRN code to be a very good candidate for the early genetic code.

General Arguments

The initial division of the nucleic acid triplets into β-turn- and β-sheet-encoding codons is a rather intriguing problem. For stability, a β-sheet requires at least two strands and thereby the folding of the peptide chain back on itself. In addition, as noted by Orgel (1977), the interstrand β-turn is in turn stabilized by the β-sheet strand pairing. Such simple structures have the potential both to form catalytically active sites in the β-turn and to function in membrane absorption via a hydrophilic–hydrophobic sidedness to the β-sheet. Indeed, a simple codon-ambiguity–reduction model of the genetic code's initial evolution (as proposed by Fitch and Upper 1986), shown in Fig. 1, can produce in only two reduction steps a code that accounts for β-turns, β-sheets, and hydrophilic–hydrophobic structures. This leaves the important, very stable α-helix structure unencoded. The strongest helix formers—Ala, Glu, Met, and Leu (Chou and Fasman 1978)—are encoded by all codon types except YRN. There is no simple way to relate this to the current proposal. However, the idea that one of the first steps in the origin of the genetic code was the development of

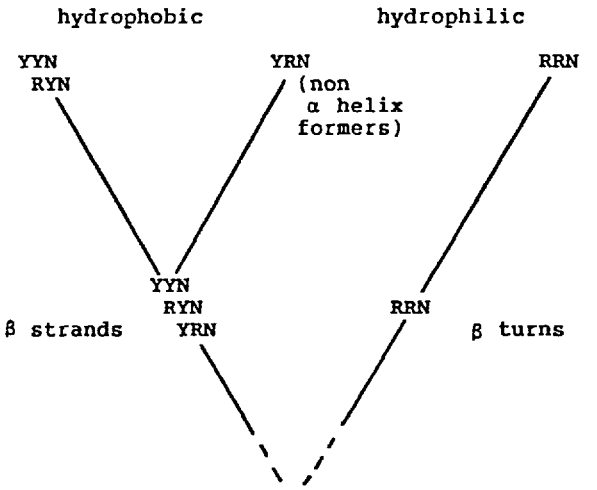

Fig. 1. Reconstruction of the evolutionary dynamics that produced the genetic code's current structure, as hypothetically induced by initial selection for β-turns. The dotted lines indicate that the root is not meaningful in an evolutionary sense, since selection can act only once there is something to select

an RRN specificity for the encoding of β-turns implies the encoding of such turns within a pool of random peptides including α-helices. The ability to encode turns immediately introduces the possibilities not only of forming β-sheets by folding the peptide back on itself but also of the termination of random helices. Thus the encoding of a β-turn provides not only for the selection of the turns based on their own intrinsic properties, but in addition for the delineation of the other two major secondary structures known.

Discussion

To be selected for, the early β-turns must have been biologically active. The idea that β-turns were the early bioactive structures is not new (Orgel 1977), although no direct experiments have been done to substantiate it. β-Turns, however, are probably associated with a variety of biological activities (Smith and Pease 1980; Rose et al. 1985). They are presumed to play an active role in immunological recognition, phosphorylation, glycosylation of proteins, and post-translational hydroxylation of proline in collagen and elastin. They also are presumed to be associated with the activity of antifreeze peptides and proteins, peptide hormones, antibiotics, toxins, antitoxins, ionophores, and metabolic products. They are frequently associated with catalytic sites in globular proteins (e.g., serine in serine proteases). From this perspective it is not very radical to propose that β-turns in short peptides might also have been carriers of useful biological activity, especially since the best β-turn formers contain active chemical groups capable of interacting not only with potential substrates but also with metals (Vogt et al. 1979). Obviously, early peptides containing β-turns were probably very non-specific and ineffective in comparison with contemporary enzymes, but they could have been sufficiently specific to maintain the internal organization of the primitive system.

Our hypothesis can be summarized as follows: (1) The abundance of Gly and formation of β-turns in the prebiotic environment might have promoted selection for optically pure forms of peptides, and ultimately elimination of one of them; (2) β-turns were the earliest biologically active protein structures selected for in the primitive organism; (3) the amino acids that are most abundant in contemporary β-turns were also preferred in those earliest β-turns; (4) these same amino acids were the first ones synthesized by the primitive metabolic pathways and the first ones encoded for by the primitive code; (5) β-turn encoding introduces the potential for encoding β-sheets as well, and perhaps for controlling the lengths of random helices.

Acknowledgments. We would like to thank Dr. Walter Fitch for his critical and useful comments, and Dr. Marek Kloczewiak for attracting our interest to the interactions of β-turns with metals. This work was done under the auspices of the MBCRR, which is supported by NIH grant RR02275-02.

References

Brack A, Orgel LE (1975) β-Structures of alternating polypeptides and their possible prebiotic significance. Nature 256: 383–387

Chou PY, Fasman GD (1978) Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol 47:45–148

Fitch WM, Upper K (1986) Evolution of the genetic code. In: Speaker abstracts. Macromolecules, genes and computers: international symposium and workshop, 12–17 Aug 1986, Waterville Valley

Harada K, Fox SW (1964) Thermal synthesis of natural amino acids from a postulated primitive terrestrial atmosphere. Nature 201:335–336

Lewis PN, Momany FA, Scheraga HA (1971) Folding of polypeptide chains in proteins: a proposed mechanism of folding. Proc Natl Acad Sci USA 68:2293–2297

Lewis PN, Momany FA, Scheraga HA (1973) Folding of polypeptide chains in proteins. Biochim Biophys Acta 303:211–229

Marlborough DI (1980) Early assignments of the genetic code dependent upon protein structure. Orig Life 10:3–14

Orgel LE (1972) A possible step in the origin of the genetic code. Isr J Chem 10:287–292

Orgel LE (1975) Prebiotic polynucleotides and polypeptides. Isr J Chem 14:11–16

Orgel LE (1977) β-Turns and the evolution of protein synthesis. In: Bradbury EM, Javaherian K (eds) The organization and expression of the eukaryotic genome: proceedings of the international symposium, 3–6 May 1976, Teheran. Academic Press, London, pp 499–504

Rose GD, Gierasch LM, Smith JA (1985) Turns in peptides and proteins. Adv Protein Chem 37:1–109

Siegel H, Martin BR (1982) Coordinating properties of the amide bond. Stability and structure of metal ion complexes of peptides and related ligands. Chem Rev 82:385–426

Smith JA, Pease LG (1980) Reverse turns in proteins. CRC Crit Rev Biochem 8:315–399

Toniolo C (1980) Intramolecularly hydrogen-bonded peptide conformations. CRC Crit Rev Biochem 9:1–44

Venkatachalam CM (1968) Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. Biopolymers 6:1425–1436

Vogt H-P, Strassburger W, Wollmer A, Fleischhauer J, Bullard B, Mercola D (1979) Calcium binding by troponin-C and homologs is correlated with the position and linear density of "β-turn forming" residues. J Theor Biol 76:297–310

von Heijne G, Blomberg C, Baltscheffsky H (1978) Early evolution of cellular electron transport: molecular models for ferredoxin–rubredoxin–flavodoxin region. Orig Life 9:27–37

Wald G (1957) The origin of optical activity. Ann NY Acad Sci 69:352–376

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

Wong TF (1975) A co-evolution theory of the genetic code. Proc Natl Acad Sci USA 72:1909–1912

Received February 16, 1986/Revised and accepted November 5, 1986