Origin of the cloverleaf shape of transfer RNA - the double-hairpin model: Implication for the role of tRNA intron and the long extra loop

Terumichi Tanaka[§] & Yo Kikuchi
Division of Bioscience and Biotechnology, Department of Ecological Engineering, Toyohashi University of Technology, Tempaku-cho, Toyohashi, Aichi 441-8580, Japan
[§] To whom correspondence should be addressed. Tel: +81-532-44-6920; Fax: +81-532-44-6929;
Email: tanakat@eco.tut.ac.jp
(Received 17 August 2001, Accepted 14 September 2001)

Abstract
By statistic analyses of tRNA sequences, we found that most tRNA sequences have vestiges of double hairpin folding. In the double hairpin folding, the acceptor- and the anticodon (and extra-) stems of tRNA are unfolded and then these unfolded regions are used to form the extended D- and T-stems, resulting in the formation of two tandemly joined stems and loops. This fact strongly suggests that structure of the tRNA molecules should be achieved through double hairpin formation in the ancient pre-biotic world. Here we show the statistic evidence of the double hairpin, and propose a double hairpin model. The double hairpin model can explain the origin of anticodon and discriminator bases of tRNA, the importance of some modified bases in tRNA, and also suggest some roles of tRNA-introns and extra loops.

Keywords: transfer RNA; ribonuclease P; ribozyme; cloverleaf structure; double-hairpin strucuture; long extra loop; tRNA intron; anticodon; discriminator

Introduction

The characteristics of living things are defined by their genes and the gene-encoded products, and also the compounds made by those products. These molecules vary in size and amount, however some molecules are identical beyond the species. Transfer RNA (tRNA) is one of such molecules. Each tRNA generally has cloverleaf structure, consisting of four short stems, three short loops, and a variable extra loop (see Fig. 1; *left*).

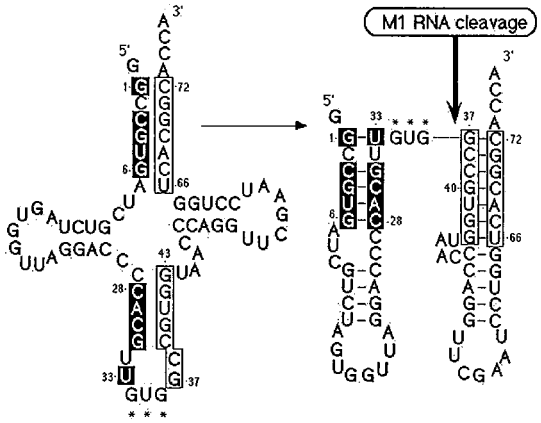


Figure 1 Schematic representation of hyperprocessing reaction of tRNA molecule by the catalytic RNA subunit of bacterial RNase P. The *Drosophila* $tRNA^{His}$ is shown in cloverleaf (*left*) and double-hairpin (*right*) folding. Bases pair-formable in the double-hairpin are shown as boxed digits. The *in vitro* cleavage site of *Drosophila* $tRNA^{His}$ by *E. coli* RNase P RNA (M1 RNA) was also indicated within an arrow [12]. The numbers for tRNA bases are according to Sprinzl *et al.* [23].

This tRNA structure is common, and the genetic codon system is also almost common beyond the species. These facts suggest the origin of the tRNA molecule appeared in the pre-biotic world, and the appearance of the tRNA is previous to the appearance of the protein synthesis system. Therefore, to reveal the origin of tRNA may be closely related to reveal the origin of life.

Where did tRNA come from? How had tRNA molecule acquired the cloverleaf structure? The tRNA structure is complicated and ordered as it was accidentally formed from random nucleotide sequences. It must come through some intermediate process. In the past decades, some models appeared to dissolve the question [1-7]. Most models lack experimental supports and do not explain the functional features of tRNA molecules. In the genomic tag model proposed by Weiner and Maizels, they argued that the tRNA-like structure functioned as a tag for ancient genomic replication. They emphasize the importance of the 'Top half' of the tRNA, made up of the acceptor stem and the T-stem/loop, because most tRNA-related enzymes recognize tRNA molecules only or mainly at this site [6,7]. Their model suggests that the 'Bottom half', made up of D-stem/loop and the anticodon stem/loop, was additionally inserted as the large bulge into the 'Top half' hairpin. Their models, however, do not explain the origin of the shape of tRNA.

How had tRNA molecule acquired the cloverleaf shape? The answer lay within the tRNA molecule itself. As according to Weiner, tRNA molecule is one of molecular fossils, which still retain the vestiges of ancient forms within the contemporary forms [7]. Some tRNAs changes their conformation from the cloverleaf to another form. In case of *Drosophila* initiator methionine tRNA, the acceptor stem and the anticodon stem were melted under *in vitro* conditions, retaining the D- and T-stems/loops, to newly form a double-hairpin folding [8-11]. The conformational changes of the tRNA molecule was experimentally confirmed using the *E. coli* ribonuclease P RNA ribozyme reaction [9,10]. The reaction was denoted as hyperprocessing. The phenomenon was not specific to this tRNA: the conformational changes of *Drosophila* alanine tRNA and

Viva Origino 29 (2001) 134- 142
© 2001 by SSOEL Japan

histidine tRNA were also detected (see ref. 12; Figure 1). These three tRNAs commonly have complementarities between the acceptor stem and the anticodon stem with extra loop region that promote to disrupt the canonical cloverleaf folding and to form unexpected double hairpin folding. The presence of unexpected complementarities in the tRNA molecules are, of course, disadvantageous to the stability of tRNA. Why do such complementarities exist? We think they are vestiges of ancient tRNAs.

In this paper, we examined the possibilities of double hairpin folding in available tRNA sequences, and statistically analyzed. Here, we propose a possible intermediate model for tRNA shape formation. Our model explains how the cloverleaf shape of tRNA molecule was acquired, and the model also suggests the strategies of the tRNA molecules to stabilize the cloverleaf shape, including possible roles of tRNA introns and the long extra loops.

Procedures for Statistics

tRNA sequences

The tRNA sequence files obtained for the seven species, *E. coli*, *B. subtilis*, *S. cerevisiae*, *A. thaliana*, *D. melanogaster*, *H. sapiens*, and *H. volcanii*, are as follows: 83 sequence files for *E. coli* tRNAs, 18 files (including files for tRNA gene cluster) for *B. subtilis*, 131 files for *S. cerevisiae*, 58 files for *A. thaliana*, 61 files for *D. melanogaster*, 88 files for *H. sapiens*, and 53 files for *H. volcanii*. Additionally, 117 files for alanine tRNAs, 95 files for glycine tRNAs, 19 files for histidine tRNAs, 51 files for initiator methionine tRNAs, 84 files for lysine tRNAs, 75 files for proline tRNAs, 120 files for serine tRNAs, 58 files for tryptophan tRNAs, and 134 files for tyrosine tRNAs, were obtained. Sequence files were obtained from the databases of EMBL and GenBank Web sites.

Statistics of tRNAs for double hairpin formation

The statistics were done using tRNA sequences. The tRNA molecule was divided to two compartments retaining the D-stem/loop or T-stem/loop as core hairpin: the 5'-half and the 3'-half. The anticodon loop was attached to both halves in the statistics. The possibility of helix-formation of between the strands N^{1}-N^{9} and N^{26}-N^{38}, and the strands N^{32}-N^{48} and N^{66}-A^{76} were examined by the procedures as illustrated in Figure 2.

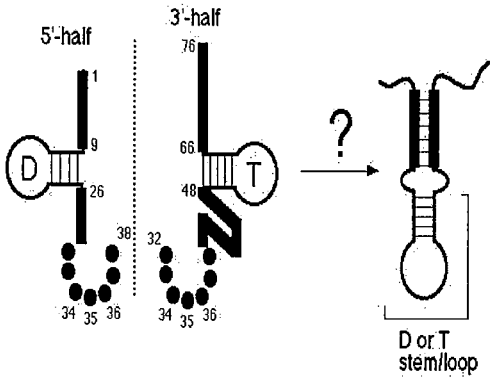


Figure 2 Schematic representation of the protocol for the statistic analysis. The tRNA molecule was divided to two compartments retaining the D-stem/loop or T-stem/loop as core hairpin: the 5'-half and the 3'-half (*left*). The anticodon loop was attached to both halves in the statistics. The possibility of helix-formation of between the strands N^{1}-N^{9} and N^{26}-N^{38}, and the strands N^{32}-N^{48} and N^{66}-A^{76} were examined.

Results and discussion

Clue to non-cloverleaf folding of tRNA

tRNA molecules have been believed to be stable and rigid for a long time. Recent studies revealed that some tRNA molecules lacking base modifications can not hold the cloverleaf structure [13,14], and some tRNA molecules move from the cloverleaf folding to the double hairpin folding retaining the D- and T-stems/loops under the *in vitro* conditions with/without the base modifications [12]. These results do not conflict with the fact that natural tRNAs are stable under *in vivo* conditions, but suggest that the organisms have acquired the systems for tRNA stabilization. The clue to the ancient tRNAs lay in the compilation of results of artificially destabilized tRNAs, because tRNA molecule itself is one of molecular fossils. As far as we know, reports on the conformational change of tRNA are rare except for above described. For example, the conformational change of *Drosophila* histidine tRNA is shown in Figure 1. Interestingly, the conformational change was done commonly using D- and T-stems/loops as core hairpins, which implies that these hairpins are stable and are the kernel of tRNA molecules. The hairpin formation of the 3'-half was experimentally confirmed [9,14]. The same type of conformational change was also observed with alanine tRNA and initiator methionine tRNA. Here, we have two questions: three *Drosophila* tRNAs above are exceptions? , or other tRNAs can fit to the double hairpin folding?

Statistics for double hairpin formation of tRNA

To answer above questions, we examined about 300 different tRNA sequences from seven species, *E. coli*, *B. subtilis*, *S. cerevisiae*, *A. thaliana*, *D. melanogaster*, *H. sapiens*, and *H. volcanii* (summarized in Table 1A). The possibility of double hairpin formation of tRNA sequences were examined retaining the D- and T-stems/loops as core hairpins as according to the examples of the *Drosophila* tRNAs, the conformation change of which was confirmed. The base pair formation between the acceptor stem bases and the anticodon stem/loop bases was examined for the 5'-half and the 3'-half molecules (see Fig. 2). The anticodon loop region (N^{32}-N^{38}) was attached to each half molecule for the statistics, to examine to which part of the molecule the anticodon bases belong.

Table 1. Summary of statistic analyses.
The tRNA sequences were obtained from the Web sites of tRNA sequence databases in EMBL and GenBank. The abbreviations, Eco, Bsu, Sce, Dme, Hsa, Hvo, and Ath, represent for *E. coli*, *B. subtilis*, *S. cerevisiae*, *D. melanogaster*, *H. sapiens*, *H. volcanii*, and *A. thaliana*, respectively.
(**A**) The tRNA sequences corresponding to Cys, Gln, Trp (of *D. melanogaster*), Cys, Asp, Ile, Trp (of *H. sapiens*), Ile, Leu, Asn, Arg, and Thr (of *A. thaliana*) were not available. Two 3'-half sequences (*E. coli* $tRNA_3^{Arg}$ and *S. cerevisiae* $tRNA_3^{Gly}$) are included. The information of the base N^{73} of four sequences (*D. melanogaster* $tRNA_3^{Ile}$, and three *S. cerevisiae* $tRNA^{Gln}$s was not available.

(A) tRNAs used for the analysis

origin	different sequences	corresponding amino acids
Eco	46	20
Bsu	36	20
Sce	51	20
Ath	25	15
Dme	36	17
Hsa	47	16
Hvo	41	20
total	282	---

(**B**) 280 tRNA sequences were used for the statics for 'D-hairpin', excepting two 3'-partial sequences (*E. coli* $tRNA_3^{Arg}$ and *S. cerevisiae* $tRNA_3^{Gly}$), and 278 tRNA sequences were analyzed for 'T-hairpin', excepting four incomplete sequences (*D. melanogaster* $tRNA_3^{Ile}$, and three *S. cerevisiae* $tRNA^{Gln}$s).

(B) Usage of anticodon three bases (N^{34}-N^{36}) for base pairing formation in 'D-, or T-hairpin'.

anticodon base	count								
	Eco	Bsu	Sce	Ath	Dme	Hsa	Hvo	total	([%])
in 'D-hairpin'									
N^{34}	23	16	27	14	9	15	17	121	(43.2)
N^{35}	11	10	8	4	3	11	10	57	(20.3)
N^{36}	7	2	4	2	2	0	5	22	(7.9)
in 'T-hairpin'									
N^{34}	1	0	3	0	7	4	3	18	(6.5)
N^{35}	3	0	2	2	4	4	7	22	(7.9)
N^{36}	6	0	6	0	4	8	11	33	(11.9)

(**C, D**) Data were analyzed using 278 tRNA sequences, excepting four sequences (*D. melanogaster* $tRNA_3^{Ile}$, and three *S. cerevisiae* $tRNA^{Gln}$s).

(C) Usage of 3'-terminal three four bases (N^{73}-A^{76}) for base pairing formation in 'T-hairpin'.

position	count								
	Eco	Bsu	Sce	Ath	Dme	Hsa	Hvo	total	([%])
N^{73}	27	17	25	16	14	26	17	142	(51.1)
C^{74}	18	9	8	3	7	22	14	81	(29.1)
C^{75}	19	10	21	4	7	13	16	90	(32.3)
A^{76}	11	7	12	13	11	16	23	93	(33.4)

(D) Details of base-pair formation at N^{73} in 'T-hairpin'.

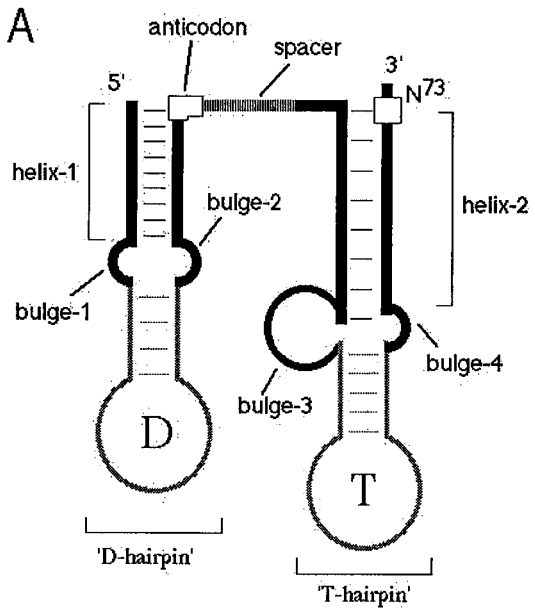
	count								
	Eco	Bsu	Sce	Ath	Dme	Hsa	Hvo	total	([%])
pair-formable in									
A^{73}-u	14	5	13	6	2	10	6	56	(20.3)
G^{73}-u	3	1	6	6	9	4	0	29	(10.4)
G^{73}-c	6	4	4	2	1	2	7	26	(9.3)
C^{73}-g	0	1	1	1	1	2	1	7	(2.5)
U^{73}-a	2	4	0	0	1	0	2	9	(3.2)
U^{73}-g	2	2	1	1	0	8	1	15	(5.4)
(sum)	27	17	25	16	14	26	17	142	(51.1)
unpaired									
A^{73}	16	13	15	6	13	17	23	103	(37.1)
G^{73}	2	5	6	3	8	4	1	29	(10.4)
C^{73}	1	1	2	0	0	0	0	4	(1.4)
U^{73}	0	0	0	0	0	0	0	0	(0)
(sum)	19	19	23	9	21	21	24	136	(48.9)

The results of the statistics are summarized in Figure 3. Many tRNA molecules did fit to the double hairpin folding (see Fig. 3A).

The acceptor stem region (mainly using N^{1}-N^{8}) and the anticodon stem/loop region (N^{26}-N^{34}) can form a helix (denoted as 'helix-1'), and the acceptor stem region (N^{66}-N^{73}) and the anticodon stem/loop with extra loop region (N^{39}-N^{47}) can also form a helix (denoted as 'helix-2'). These helices locate near or on the retained D- or T-stem/loop. Two hairpins, denoted as D-hairpin and T-hairpin, are joined by a spacer region. The possible base pair formation in the helix-1 and helix-2 are shown in Figure 3B.

Surprisingly, not a little tRNAs did fit to the double hairpin folding. Among them, $tRNA^{Leu}$, $tRNA^{Ser}$, and $tRNA^{Tyr}$ are commonly of high count for base pair formation, in both helix-1 and helix-2. These data indicated that the double hairpin folding is not specific to three *Drosophila* tRNAs but adaptable to bacterial, eukaryotic, and archaeal tRNAs. The results also showed that the average counts of base pair formation were larger than the theoretical values, and that indicates many contemporary tRNA molecules have vestiges of double hairpin folding.

Figure 3C shows the base usage in base pair formation in helix-1 or helix-2. The regions N^{1}-N^{8} and N^{26}-N^{34} were mainly used for helix-1, and the regions N^{39}-N^{47} and N^{66}-N^{73} were for helix-2. Interestingly, the results showed that the anticodon bases, at least the base N^{34}, belonged to the 5'-half tRNA molecule (the D-hairpin), and also the base N^{73} belonged to the edge of the 3'-half (the T-hairpin). The region around N^{37} and N^{38}, the point of which the tRNA intron is sometimes inserted, was included in neither hairpin. These results suggested the possibility that the wobbling codon third letter and the discriminator base are the double hairpin-originated. The details of the statistics for N^{34} and N^{73} are summarized in Table 1B,C,D.



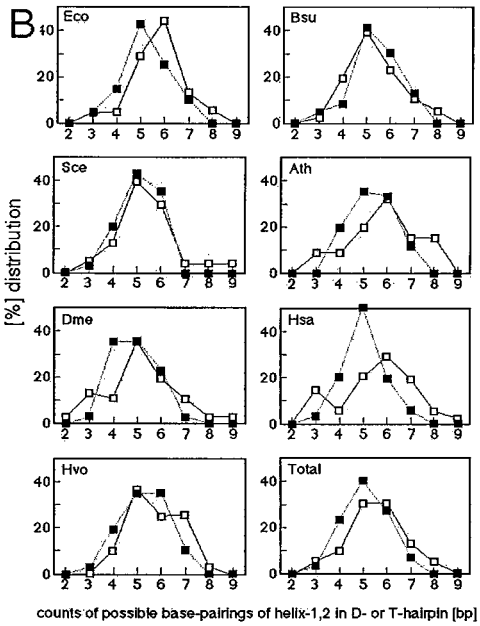
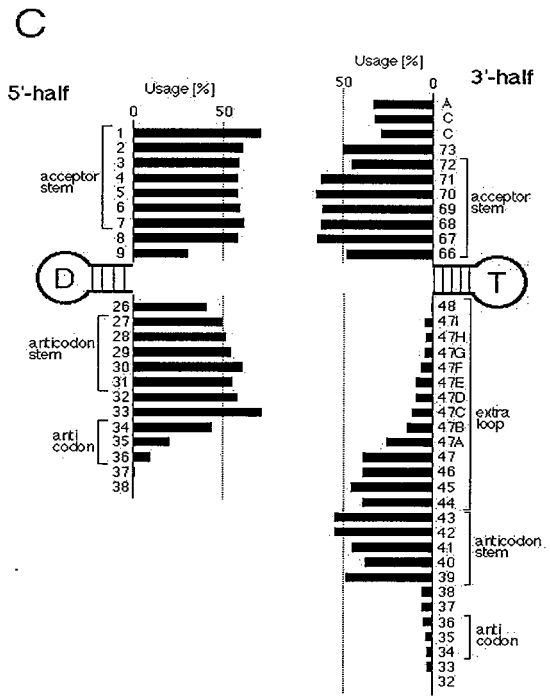


Figure 3 Summary of the double hairpin folding from tRNA sequences.

(A)A schematic representation of 'double hairpin' folding of tRNA molecule. The acceptor stem and the anticodon stem can be destroyed and two newly helices, helix-1 and helix-2, can be formed. The maximum length of the helix can be 9 for the helix-1 and 11 for the helix-2 (see Fig. 2). The 5'- and the 3'-half molecules form 'D-hairpin' and 'T-hairpin', retaining the D-stem/loop and T-stem/loop, respectively. The D-hairpin includes small bulge(s), bulge-1 and/or bulge-2. The T-hairpin has bulge(s), bulge-3 and/or bulge-4. The bulge-3 is large in the cases of tRNA with a large extra loop, such as $tRNA^{Leu}$ and $tRNA^{Ser}$. The new helices formed in the double hairpin are indicated as thick lines. The retained D- and T-stems/loops are indicated as gray. Two hairpins are connected with a spacer sequence between them. In most cases, the anticodon three bases locate adjacent to, at the edge of, or within the 'D-hairpin'.

(B)Distribution of the count of possible base pairings of the helix-1 in 'D-hairpin' (closed squares) and the helix-2 in 'T-hairpin' (open squares) for tRNAs from two bacteria, a yeast, a plant, two animals, an archaea, and sum of these seven species. The averages of the pairing count for the helix-1 in 'D-hairpin' are 5.13 (*E. coli*), 5.42 (*B. subtilis*), 5.16 (*S. cerevisiae*), 5.33 (*A. thaliana*), 4.83 (*D. melanogaster*), 4.94 (*H. sapiens*), 5.29 (*H. volcanii*), and 5.14 (seven species), respectively. The value of expected base pairings between random nine-base strands is about 3.38 calculated from the term $S(3/8)^{i}(5/8)^{9-i}{}_{9}C_{i}i$. The averages of the pairing count for the helix-2 in 'T-hairpin' are 5.69 (*E. coli*), 5.36 (*B. subtilis*), 5.38 (*S. cerevisiae*), 5.78 (*A. thaliana*), 5.06 (*D. melanogaster*), 5.57 (*H. sapiens*), 5.75 (*H. valcanii*), and 5.51 (seven species), respectively. The value of expected base pairings between random 11-base strands is about 4.13 calculated from the term $S(3/8)^{i}(5/8)^{11-i}{}_{11}C_{i}i$.



(C) Usage of bases in base-pair formation for the 'D-hairpin' (*left*) and the 'T-hairpin' (*right*) in the double hairpin folding. The degree of usage is expressed by length of bars. The helix-1 is mainly formed by the regions N^{1}-N^{8} and N^{26}-N^{34}, and the helix-2 is mainly formed by the regions N^{39}-N^{47}andN^{66}-N^{73}.

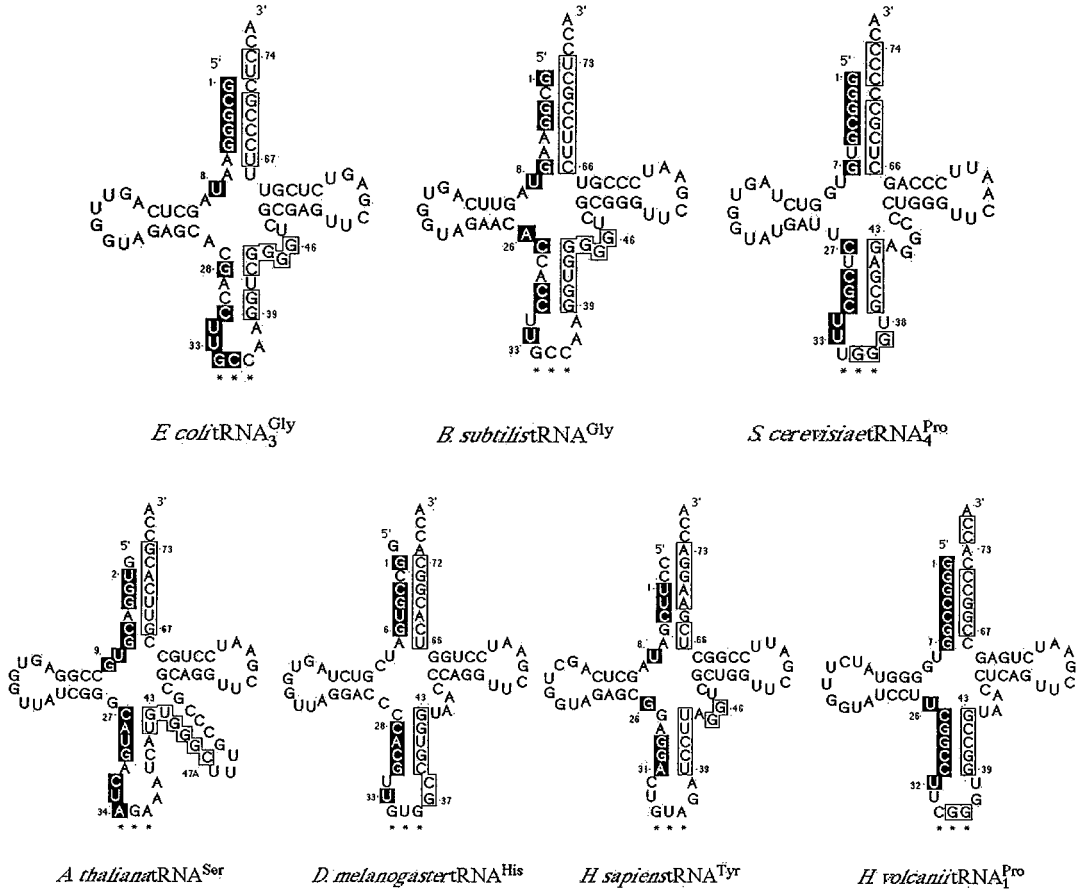


Figure 4 Examples of 'double hairpin' formable tRNAs from the cloverleaf structure. tRNAs are shown in the cloverleaf folding. The anticodon bases are indicated with asterisks. Bases pair-formable in the double-hairpin are shown as boxed digits. Base pairing formation is predicted based on A-U, G-C, and G-U pairs. The nucleotide nomenclature was done according to Sprinzl *et al.* [23].

In about 43% of the tRNAs, the base N^{34} was base-paired in D-hairpin, and in about 51% of tRNAs, the base N^{73} was also base-paired in T-hairpin. The details of N^{73} showed interesting results that most of pyrimidine bases at N^{73} were included in T-hairpin formation. The results also suggested the origin of the 3'-terminal bases, C^{74}, C^{75}, and A^{76}. Compared to other bases, the 3'-terminal CCA bases were not frequently included in the T-hairpin formation: only about 30% were used (Table 1C; Fig. 3C).
The result was consistent with the fact that many tRNA sequences lack this CCA sequence in their genomic sequences. These facts suggest that the CCA sequence of tRNA was added, in many cases, to the ancient template tRNAs following to acquisition of the cloverleaf shape. Examples of tRNAs, which had high counts for base-pair formation in the double hairpin folding, are shown in Figure 4. Please note that the base N^{34} was used for base-pairing in helix-1 in the double hairpin folding in cases of *E. coli* glycine tRNA, and *A. thaliana* serine tRNA, and also the base N^{73} was used for base-pairing in helix-2 in the double hairpin folding in cases of *E. coli* glycine tRNA, *B. subtilis* glycine tRNA, *S. cerevisiae* proline tRNA, *A. thaliana* serine tRNA, and *H. sapiens* tyrosine tRNA. When the anticodon bases include guanine(s), they can be base-paired with the 3'-terminal C^{74} or/and C^{75} (see Fig. 4); but this type base pair formation was still minor in the statistics (see Fig. 3C; Table 1B,C).

For the next step of statistics, we also examined the double hairpin formability of tRNAs from many available species: 71 $tRNA^{Ala}$s (58 species), 58 $tRNA^{Gly}$s (35 species), 9 $tRNA^{His}$s (9 species), 61 $tRNA^{Lys}$s (35 species), 37 $tRNA^{iMet}$s (33 species), 62 $tRNA^{Pro}$s (39 species), 89 $tRNA^{Ser}$s (34 species), 30 $tRNA^{Trp}$s (24 species), and 61 $tRNA^{Tyr}$s (33 species), were used for the statistics. The statistics indicated that these tRNAs commonly could also be adapted to the double hairpin folding as above tRNAs from the seven species (data not shown). The results also showed that the count for the helix-formation in the helix-1 or helix-2 was especially high in case of $tRNA^{Ser}$s and $tRNA^{Tyr}$s: the common feature was that these tRNAs have long extra loop or tRNA intron. The count for the helix-2 formation was also commonly high in cases of metazoan initiator methionine tRNAs and metazoan lysine tRNAs.

The statistics showed that there was no correlation between the helix-1 formation and the helix-2 formation; two helices were statistically independent. The contemporary tRNAs can be divided to four groups: (i) tRNAs highly fit to the double hairpin folding, (ii) tRNAs of which the 3'-halves fit to the hairpin, but the 5'-halves do not fit, (iii) tRNAs of which the 5'-halves fit to the hairpin, but the 3'-halves do not fit, and (iv) tRNAs that do not fit to double hairpin folding. The group-i contains long extra loop-containing tRNAs such as serine and leucine tRNAs, and intron-containing tRNAs such as tyrosine tRNAs. The tRNAs shown in Figure 4 also belong to this group. The group-ii contains metazoan initiator methionine and lysine tRNAs. Interestingly, the primer tRNAs for viral reverse transcription belong to the group-i or group-ii [8,15]. The fact that the primer tRNAs considered as molecular fossils had high counts for the double hairpin folding will support our double hairpin hypothesis. The group-iii contains archaeal lysine tRNAs. The group-iv contains eukaryotic valine tRNAs. Generally, when many isoacceptors exist for a cognate amino acid residue and they differ in nucleotide sequences, the double hairpin formability of tRNAs is low: that suggests that to posses many different tRNA isoacceptors is one of strategies for tRNA molecules to evolve and to achieve effective protein synthesis system.

Surprisingly, the double hairpin model could even be adapted to non-canonical cloverleaf tRNA: *A. thaliana* $tRNA^{Gln}$, of which the D-stem/loop is short, also did fit to the double hairpin folding (data not shown).

Above results strongly suggest that the double hairpin folded tRNA was the direct origin, or the intermediate for the cloverleaf formation process.

Double hairpin intermediate model

The above statistics for tRNA sequences showed that the contemporary tRNA molecules have many vestiges of double hairpin folding within their RNA sequences, that strongly suggests that tRNA molecules have been acquired through double hairpin intermediate in the shape formation process. Here, we propose a double hairpin intermediate model for tRNA shape formation (see Fig. 5).
At the beginning, in the pre-biotic RNA world, many RNA hairpins existed here and there. Some of them contained CCA sequence. Some hairpins were found as continuously joined hairpin clusters, or some hairpins were found as independent forms. The shuffling of these hairpins might have existed by ancient catalytic RNAs which cleaved and ligated such RNA hairpins in *cis* and in *trans*, like the contemporarily found and obtained catalytic RNAs by *in vitro* selections [16-21]. Among them, there might be two hairpins which contained some complementary sequences enough to form a cloverleaf-like structure. Some RNA was experimentally confirmed that RNA folding can be swithced and alternated by only one or two base(s) replacement [22]. The spacer region between two hairpins was takes into at the corresponding site of anticodon loop and/or extra loop. In some cases, the moieties of hairpin were taken into a helix of long extra loop and then were stabilized in the cloverleaf-like shape. In some cases, the residual flanking region at the anticodon loop region formed an additional mini-helix, which also stabilized the cloverleaf-like shape. And then, the template

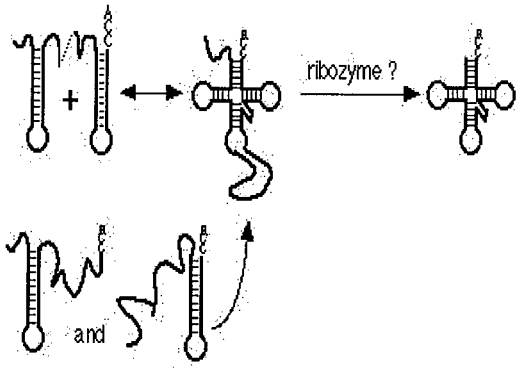


Figure 5 A schematic representation of the cloverleaf shape formation of tRNA molecule in the double hairpin intermediate model. In the prebiotic world, many RNA hairpins, the simplest RNA structure, containing some base mismatches, existed. These hairpins might be shuffled randomly to form newly hairpin clusters. In some combinations, the proximal regions of two clustered RNA hairpins had complementarity (*above left*), enough to form a cloverleaf or a cloverleaf-like folding (*above middle*). The flanking region between two hairpins was taken into the cloverleaf folding at the anticodon loop region or the extra loop region, which might be origin of the tRNA intron or the extra loop. After the birth of the cloverleaf shape, some template RNAs were selected and chosen for the ancient tRNAs. Once the tRNA creating system had made, non-double hairpin RNAs, made up of one hairpin and one flexible region (*below left*), were recruited additionally as candidates for the ancient tRNAs.

RNAs in the cloverleaf-like shape, the candidates for the ancient tRNAs, were exposed to the selections by some ancient tRNA-related enzymes such as the ancient RNA component of ribonuclease P. We think that the RNA component of ribonuclease P is also one of molecular fossils defined by Weiner [7], therefore we cannot neglect the possibility of coevolution of tRNA molecules and ribonuclease P. Once the selection system of cloverleaf shape was established, other candidate RNAs could be easily recruited from the pools of RNA hairpins.

The hypothesis of the tRNA shape formation process is simply summarized that tRNA molecules are formed mainly by joining two hairpins, not inserting functional domains into the 'Top half' of tRNA formed by the acceptor stem and the T-stem/loop.

Roles of tRNA intron and long extra loop

The double hairpin intermediate model explains the pathway of formation of the cloverleaf shape tRNA from two RNA hairpins. As the cloverleaf shape is required for the tRNA function, to be recognized by many tRNA-related enzymes, the cloverleaf shape has to be stabilized by some mechanism, otherwise the conformation of the molecule shall be changed and might be degraded by some enzymes such as ribonuclease P as shown in Figure 1. Interestingly, the statistic results showed that the vestiges of the double hairpin highly remain in or are retained by tRNAs which have long extra loop or tRNAs having intron at the anticodon loop. A plant $tRNA^{Ser}$ and a human $tRNA^{Tyr}$ are shown in Figure 5 as examples. In *Arabidopsis* $tRNA^{Ser}$, the regions of U^{7}-G^{9} and C^{27}-A^{34}, and also the regions U^{42}-C^{47A} and G^{67}-G^{73} have complementarity to form the double hairpin folding (Fig. 6A).

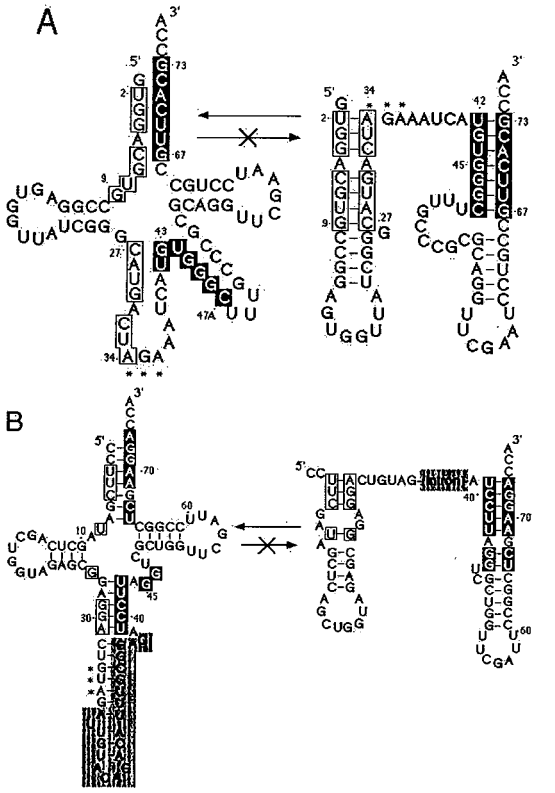


Figure 6 tRNAs which contain long extra loop or intron sequence at the anticodon loop.
(A) *Arabidopsis* $tRNA^{Ser}$.
(B) Human $tRNA^{Tyr}$ with the intron sequence. tRNAs are shown in cloverleaf folding (*left*) and in double hairpin folding (*right*). The bases contributing to form double hairpin structure are boxed. The intron sequence for human $tRNA^{Tyr}$ is shadowed.

In human $tRNA^{Tyr}$, the regions U^{3}-U^{9} and G^{26}-A^{31}, and also the regions U^{39}-G^{46} and U^{66}-A^{73} have complementarity to form the double hairpin folding (Fig. 6B). In both cases, the extra sequences contribute to form additional mini-helix or stem that seem to stabilize and fix the tRNA to the cloverleaf folding. The same tendency was found with other tRNAs, which have long extra loop or intron sequence. The fact that the base modifications of tRNA, one of stabilizers for the cloverleaf shape, are done previous to the cleavage of tRNA intron, is not contradictory to but consistent with above hypothesis: the role of tRNA intron, to stabilize the cloverleaf folding, is, then communicated to the modified bases.

In our double hairpin intermediate model, the tRNA molecules have evolved from the double

hairpin intermediate to the cloverleaf folding, stabilizing the cloverleaf shape by base modifications, inserting long extra loops, retaining introns, and base replacement of the key positins of tRNA.

Conclusion

The statistic results of tRNA sequences showed that there are many vestiges of double hairpin intermediate process within the tRNA molecules, the evidence for the double hairpin model. We think that tRNA molecules are still on the process of evolution from the ancient forms to more stable and efficient forms, compiling base replacements, base modifications, and obtaining stabilizers such as helical formed intron or long extra loop. As the tRNA molecules are deeply concerned with many biological phenomena, to reveal the origin of tRNA and to reveal the stabilizing mechanism of tRNA molecules will give a clue to answer to many RNA-related unsolved questions.

Acknowledgements

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science, Sports and Culture of Japan, and in part by a Grant from "Research for the Future" Program of the Japan Society for the Promotion of Science (JSPS-RFTF97I00301).

References

1. Bloch, D. P., McArthur, B., and Mirrop, S. tRNA-rRNA sequence homologies: Evidence for an ancient modular format shared by tRNAs and rRNAs. BioSystems 17, 209-225 (1985).
2. Dick, T. P., and Schamel, W. W. A. Molecular evolution of transfer RNA from two precursor hairpins: Implication for the origin of protein synthesis. J. Mol. Evol. 41, 1-9 (1995).
3. Giulio, M. D. On the origin of the transfer RNA molecule. J. Theor. Biol. 159, 199-214 (1992).
4. Giulio, M. D. Was it an ancient gene codifying for a hairpin RNA that, by means of direct duplication, gave rise to the primitive tRNA molecule? J. Theor. Biol. 177, 95-101 (1995).
5. Giulio, M. D. The non-monophyletic origin of the tRNA molecule. J. Theor. Biol. 197, 403-414 (1999).
6. Maizels, N., and Weiner, A. M. in The RNA World. Gesteland, R. F., and Atkins, J. F., ed. (Cold Spring Harbor Lab., Cold Spring Harbor, N.Y.) pp. 577-602 (1993).
7. Weiner, A. M., and Maizels, N. 3' Terminal tRNA-like structures tag genomic RNA molecules for replication: Implication for the origin of protein synthesis. Proc. Natl. Acad. Sci. USA 84, 7383-7387 (1987).
8. Kikuchi, Y., Ando, Y., and Shiba, T. Unusual priming mechanism of RNA-directed DNA systhesis in *copia* retrovirus-like particles of *Drosophila*. Nature 323, 824-826 (1986).
9. Kikuchi, Y., Sasaki, N., and Ando-Yamagami, Y. Cleavage of tRNA within the mature tRNA sequence by the catalytic RNA of RNase P: Implication for the formation of the primer tRNA fragment for reverse transcription in *copia* retrovirus-like particles. Proc. Natl. Acad. Sci. USA 87, 8105-8109 (1990).
10. Kikuchi, Y., and Sasaki, N. Hyperprocessing of tRNA by the catalytic RNA of RNase P. J. Biol. Chem. 267, 11972-11976 (1992).
11. Kikuchi, Y. RNase P as hyperprocessing enzyme: A model for formation of a biologically functional tRNA fragment. Mol. Biol. Rep. 22, 171-175 (1996).
12. Hori, Y., Baba, H., Ueda, R., Tanaka, T., and Kikuchi, Y. *In vitro* hyperprocessing of *Drosophila* tRNAs by the catalytic RNA of RNase P: The cloverleaf structure of tRNA is not always stable? Eur. J. Biochem. 267, 4781-4788 (2000).
13. Helm, M., Brúle, H., Degoul, F., Cepanec, C., Leroux, J.-P., Giegé, R., and Florentz, C. The presence of modified nucleotides is required for cloverleaf folding of a human mitochondrial tRNA. Nucleic Acids Res. 26, 1636-1643 (1998).
14. Helm, M., Giegé, R., and Florentz, C. A Watson-Crick base-pair-disrupting methyl group (m^1A^9) is sufficient for cloverleaf folding of human mitochomdrial $tRNA^{Lys}$. Biochemistry 38, 13338-13346 (1999).
15. Mak, J., and Kleiman, L. Primer tRNAs for reverse transcription. J. Virol. 71, 8087-8095 (1997).
16. Altman, S., Kirsebom, L. A., and Talbot, S. in tRNA structure, biosynthesis, and function: Recent studies of RNase P. Söll, D., and RajBhandary, U. L., ed. (American Society for Microbiology Press. Washington, DC), pp.67-78, (1995).
17. Bartel, D. P., and Szostak, J. W. Isolation of new ribozymes from a large amount pool of random sequences. Science 261, 1411-1418 (1993).
18. Bartal, D. P., and Unrau, P. J. RNA-catalyzed nucleotide synthesis. Nature 395, 260-263 (1998).
19. Ekland, E. H., and Bartel, D. P. RNA-catalyzed RNA polrimerization using mucleotide triphosphates. Nature 382, 373-376 (1996).
20. Haseloff, J., and Gerlach, W. L. Simple RNA enzymes with new highly specific endoribonuclease activities. Nature 334, 585-591 (1988).

21. Zang A. J., Been, M. D., and Cech, T. R. The *Tetrahymena* ribozyme acts like an RNA restriction endonuclease. Nature 324, 429-433 (1986).
22. Schultes, E. A., and Bartel, D. P. One sequence, two ribozymes: Implication for the emergence of new ribozyme folds. Science 289, 448-452 (2000).
23. Sprinzl, M., Moll, J., Meissner, F., and Hartmann, T. (1985). Compilation of tRNA sequences. Nucleic Acids Res. 13, Suppl., r1-r41 (1985).