Degeneracy of the genetic code and stability of the base pair at the second position of the anticodon

JEAN LEHMANN and ALBERT LIBCHABER
Center for Studies in Physics and Biology, The Rockefeller University, New York, New York 10065, USA

ABSTRACT
With an analysis of the structural constraints of the anticodon–codon interaction within the decoding center of the ribosome, we show that the extent of degeneracy at the third position of the anticodon is determined by the level of stability of the base pair at the second position.

Keywords: tRNA; anticodon loop; U-turn; decoding center; wobble; degeneracy

INTRODUCTION
In 1978, Lagerkvist noticed that the level of degeneracy of each of the 16 groups of codons as specified by the two first codonic positions is determined by only three parameters (Lagerkvist 1978). When the two first positions form “strong” Watson–Crick base pairs with the anticodon (SS; S = G or C), the corresponding families are all fourfold degenerate; when these two positions form “weak” Watson–Crick base pairs (WW; W = A or U), the corresponding families are all twofold degenerate. The category of degeneracy of the intermediate cases (WS and SW) is determined by the purine/pyrimidine (R/Y) type of base at the second codonic position: A pyrimidine specifies a fourfold degenerate codon family whereas a purine indicates that the group of codons is split into two twofold degenerate codon families.

Lagerkvist suggested the following interpretation: “codons may be read according to the two-out-of-three principle that relies mainly on the Watson–Crick base pairs, while the mispaired nucleotides in the third codon and anticodon wobble positions make a comparatively small contribution to the total stability of the reading interaction” (Lagerkvist 1986).

The aim of the present work is to provide a structural interpretation of Lagerkvist’s parameters. Our analysis is based on the architecture of the tRNA anticodon loop (Auffinger and Westhof 1999, 2001) and on the organization of the decoding center of the ribosome (Ogle et al. 2001, 2002). We point out that the level of stability of the base pair at the second position of the anticodon plays a key role for the degeneracy. Our results show that the perturbation of this base pair by the wobble interaction can be sensed by the ribosome and constitutes the main source of selectivity for the base pair at the third position.

RESULTS AND DISCUSSION
Degeneracy in the genetic code and interpretation of Lagerkvist’s parameters
We address the rational behind the existence of the two main codon degeneracy families in the genetic code table (Fig. 1A). These two families are sufficient to characterize the symmetrical codes of the mitochondria and some primitive bacteria (Lehmann 2000).

The basic assumption on which the model is built is that wobbling was initially maximized. This maximization enables a reduction of the set of tRNAs necessary to translate small genomes. Why then are there not only fourfold degenerate codon families in the table? We show that the answer lies in the structural interpretation of Lagerkvist’s parameters.

The first two parameters (strength of N36–N1 and N35–N2 base-pair associations) are self-explanatory. The structural interpretation of the third parameter (second position purine or pyrimidine) has not been clarified yet; we discuss it below.
In order to accommodate the decoding center of the ribosome, the anticodon loop of all tRNAs has to bear some standard features (Quigley and Rich 1976; Dix et al. 1986; von Ahsen et al. 1997; Sundaram et al. 2000; Stuart et al. 2003). The main characteristic of the loop relies on the presence of the U-turn motif (Quigley and Rich 1976), which requires several conserved interactions between residues of the loop. Among these, a critical hydrogen bonding pattern involves the 2’ OH group of U33 and the base at position 35. The result depends on whether N35 is a purine or a pyrimidine (Auffinger and Westhof 2001).

A purine enables the formation of a conventional donor–acceptor hydrogen bond between (U33)O2’H and (R35)N7, bifurcated to (G35)O6 in the case of a guanine. The distance and angle are compatible with the formation of a strong hydrogen bond, i.e., $\zeta 1.3 \text{ kcal mol}^{-1}$ (Jiang and Lai 2002).

With a pyrimidine at position 35, the interaction is much weaker since the approximate position of the (R35)N7 atom is occupied by a (Y35)C5-H group, which potentially only allows for the formation of a CH···O type of hydrogen bond, with (U33)O2’. The distance of 3.5 Å estimated from MD simulations is compatible with a hydrogen bond of $\sim 0.4 \text{ kcal mol}^{-1}$ (Jiang and Lai 2002).

Based on this analysis, the third parameter is defined as follows: We credit U33–N35 with S (“strong”) when N35 is a purine and with W (“weak”) when N35 is a pyrimidine.

The structural significance of the three parameters can now be examined in Figure 1B,C. Figure 1B shows that the sum of three structural energies determines whether a codon belongs to a twofold or a fourfold degenerate family while Figure 1C shows that each of these energies contributes to the level of stability of the base pair at the second position, N35–N2:

- The first parameter (1) indicates the extent up to which N35–N2 is stabilized by a nearest neighbor, N36–N1 (Mathews et al. 1999).
- The second parameter (2) indicates the intrinsic N35–N2 base-pair stability.
- The third parameter (3) indicates whether N35 is further stabilized by U33.

The connection between the three parameters and the type of degeneracy allowed at the third position (N34–N3) requires an examination of the anticodon–codon interaction within the decoding center. Three residues of the ribosome (G530, A1492, and A1493, *Escherichia coli* numbering) interact with the minor groove of the anticodon–codon complex (Ogle et al. 2001). A1493 spans over the minor groove of the first base pair while A1492 and G530 span over the second base pair (Fig. 1C). When the base pair at the first position is G•C or C•G, the A-minor interaction realized by A1493 makes this base pair even

![FIGURE 1.](image-url)
stronger (V. Ramakrishnan, pers. comm.). It may explain why the contributions of the two first parameters to the stability of the base pair at the second position are identical (Fig. 1B).

Both N36–N1 and N35–N2 base pairs have to bear the Watson–Crick geometry to enable the optimal hydrogen-bonded interactions with the ribosome residues. Crystal structures however show that this type of monitoring does not occur at the third position: only one direct hydrogen bond is observed, between (G530)O6 and the 2′OH of the ribose of N3 (Ogle et al. 2001, 2002).

The fact that the ribosome does not directly check the base pair at the third position implies that it can only be sensed by its nearest neighbor, N35–N2.

Considering U, C, G, and A, three types of N34–N3 pairings are possible: Watson–Crick (WC), Wobble (Wb) (G+U or U+G), and base–base mismatch (Mm). Because the width of the base–base mismatches is larger or smaller than a regular base pair, they impose a stronger constraint on N35–N2 by distorting the double helix (Agris et al. 2007). In a crude approximation, N34–N3 Watson–Crick, and Wobble interactions are credited with S (‘‘stabilizing’’), while the base–base mismatches are credited with S-p, p standing for a mismatch penalty.

All local structural constraints in discussing the stability of the Watson–Crick geometry of N35–N2 have now been introduced. This parameter is critical since it enables the productive conformation of the decoding center of the ribosome (Ogle et al. 2002).

Table 1 shows that, below a total number of six consolidating hydrogen bonds defined by Lagerkvist’s parameters, the base–base mismatch at the third position sufficiently destabilizes N35–N2 to prevent translation. It implies that two tRNAs are required for the translation of the four codons, which belong to two contiguous twofold degenerate codon families. With six and seven hydrogen bonds, a base at position N34 of a tRNA (often a U in mitochondria) is able to base pair with U, C, G, or A on the codon, and none of these interactions sufficiently perturb N35–N2 to prevent translation. These codons therefore belong to a fourfold degenerate codon family.

The twofold → fourfold transition establishes a threshold for translation at 6 + S-p (in hydrogen bonds units), which can be compared to 4 + S, the weakest interaction still allowing translation. While assuming that both are roughly equal, one gets p = 2 hydrogen bonds, which compares well with known experimental values (Mathews et al. 1999).

Effects of base modifications

Although the analysis shows that the organization of the degeneracy can be described without taking base modifications into account, these are necessary to fully explain two effects.

N34 modifications play a significant role in the accuracy of wobbling. In the simplest systems (such as the mitochondria), none of the tRNAs assigned to the fourfold degenerate codon families requires a N34 modification in order to wobble over the four bases (Suzuki 2005). In twofold degenerate codon families, however, N34 modifications (e.g., xm5s2U derivatives) are always present and have been shown to prevent ‘‘leaking’’ wobblings between contiguous families (Yokoyama et al. 1985; Yokoyama and Nishimura 1995). Thus, a sharp demarcation between the twofold and fourfold degenerate codon families could not be established solely on the basis of one hydrogen bond (Table 1).

Some codons cannot be translated if specific base modifications are not present on their cognate tRNAs (Takai et al. 1999; Takai and Yokoyama 2003; Agris et al. 2004; Agris et al. 2007). For example, N37 modifications (often r4A or m1G) compensate for weak N36–N1 base pairs, in particular by increasing stacking (Grosjean et al. 1998; Konevega et al. 2004). The r6A37 modification has also been shown to restore the otherwise absent canonical U-turn in some tRNAs (Sundaram et al. 2000).

In brief, base modifications remove coding ambiguities in contiguous twofold degenerate codon families and shape unstructured tRNA anticodon loops for translation.

Table 1 represents a situation found in some primitive bacteria and mitochondria, where N34–N3 base-pairing rules and degeneracy families match exactly (Bonitz et al. 1980; Inagaki et al. 1995; Suzuki 2005), meaning that the number of different tRNAs necessary for translation is at the minimum. As far as evolved genetic systems are concerned, some structural constraints require that more than one tRNA is usually necessary for the translation of all the codons in many codon families (Agris et al. 2007; Näsälv et al. 2007).

Supports to the model

The third of Lagerkvist’s parameters, which we showed to be connected with the stability of N35 with respect to U33, is not directly involved in the anticodon–codon interaction. One could then ask whether this stabilization indeed affects translation. The earliest investigations on the U-turn already showed that it is the case (Uhlenbeck et al. 1982; Dix et al. 1986). For instance, direct binding experiments of tRNAs on P-sites programmed with complementary codons show that when U33 is replaced with dU33 in E. coli tRNA Phe (an operation that removes the critical 2′OH), the dissociation constant undergoes a 50-fold increase (Von Ahsen et al. 1997). A control experiment shows that the same substitution on tRNA Asn has only a minor effect on the corresponding dissociation constant. Since tRNA Asn has a U at position 35, no (strong) hydrogen bond is originally present, meaning that no effect is expected from the substitution.
<table>
<thead>
<tr>
<th>No. hydrogen bonds</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>N34–N3 pairing</td>
<td>WC</td>
<td>Wb</td>
<td>Mm</td>
<td>Mm</td>
</tr>
<tr>
<td>Translation</td>
<td>T*</td>
<td>T*</td>
<td>-*</td>
<td>-*</td>
</tr>
</tbody>
</table>

This table refers to the simplest genetic systems known (Mycoplasma, mitochondria). (a) Number of stabilizing hydrogen bonds defined by Lagerkvist’s parameters. (b) tRNA1: Type of base pair at the third position of the codon. (WC) Watson–Crick; (Wb) Wobble; (Mm) mismatch. This base pair affects the stability of N35–N2: (S) stabilization; (p) penalty. (c) tRNA2: Type of base pair at the third position of the codon. (WC) Watson–Crick; (Wb) Wobble; (Mm) mismatch. This base pair affects the stability of N35–N2: (S) stabilization; (p) penalty. (T) translation; (-) no translation; (T*, -*) Accuracy of decoding usually conditioned by N34 modification(s).
Another type of data related to the N35 nucleotide supports our analysis. In a very small number of tRNAs, this position is occupied by pseudouridine (ψ). This base is similar to uridine, but the critical C5–H is replaced with N1–H, which allows for a stronger hydrogen bond with (U33)O2′ (Auffinger and Westhof 2001). In addition, RNA melting experiments have shown that the ψ–A base pair itself is stronger than U–A (Ward and Reich 1968). Our model therefore predicts a ψ35 → U35 substitution on a tRNA to correspond to (one to two) S → W substitutions on the diagram of Figure 1B, meaning a possible switch from a four-way wobble to a two-way wobble. This effect has been documented in the case of a plant cytoplasmic tRNA^{Tyr}, which has an anticodon 3′GψA′ (Zerfass and Beier 1992). In the specific context of a tobacco mosaic virus (TMV) reading frame, this tRNA is able to suppress both UAG and UAA stop codons (which add to the regular translation of UAU and UAC).

In vitro translation experiments showed that the single U35 mutation completely abolishes the suppression activity.

**Influence of the N32–N38 base pair**

Another fundamental feature of the anticodon loop, the N32–N38 closing base pair (see Fig. 1C), has also been shown to influence the extent of wobbling at the third position (Lustig et al. 1993). At the level of the gene, this base pair is usually C•A (48%), T•A (18%), T•T (11%), C•C (8%), or T•C (8%). With the exception of T•A, which can form a U•A Watson–Crick base pair in the tRNA (as observed in the unbound state), crystallographic data show that the other mentioned couples are usually connected with a single bifurcated hydrogen bond (Auffinger and Westhof 1999).

In vitro translation experiments have shown that if the U32–A38 closing base pair of E. coli tRNA^{Gly} (with a modified anticodon 3′GψA′) is converted into C•A, the wobbling ability of this tRNA is extended from GGR to GGN (N = U, C, G, or A) (Lustig et al. 1993). Another study showed that the dissociation constant of E. coli tRNA^{Ala} bound to a programmed A-site is reduced by about one order of magnitude if the (rare) A32–U38 wild type is converted into U•U, A•A, or U•A, but it remains high if it is converted into CG (Olejniczak et al. 2005).

The above examples show that the stability of the anticodon–codon interaction within the decoding center can be impaired when N32–N38 may form a Watson–Crick type of interaction (Olejniczak et al. 2005). In this case, the loop does indeed not adopt the (mandatory) U-turn geometry (Auffinger and Westhof 1999). With regard to the observed effect on the wobble (Lustig et al. 1993), a simple explanation is that the U32–A38 base pair adopts a Watson–Crick type of interaction in the free state. Upon codon binding, the U-turn geometry is restored only if the anticodon–codon interaction is strong enough to force the U32–A38 Watson–Crick to switch into the single hydrogen-bonded interaction. This would make any further penalization by an N34–N3 base–base mismatch prohibitive.

The fact that a N32–N38 parameter is not required to explain the organization of the degeneracy families in the genetic code table is consistent with our hypothesis of wobble maximization at the origin. It only implies that the N32–N38 base pair has to be non-Watson–Crick.

**CONCLUSION**

Our present analysis shows that the organization of the degeneracy families in the genetic code table reflects the main structural constraints of the anticodon–codon association within the ribosome decoding center. In particular, it points out an implication of the canonical U-turn of the tRNA anticodon loop, a motif that contributes to the degeneracy at the third position. It identifies a model for the selection of the wobble base pair by the ribosome and shows that the core of the wobble rules is established independently of base modifications.

**ACKNOWLEDGMENTS**

This work originated from a discussion with V. Ramakrishnan. We thank C. Köhrer, U. Rajbhandary, V. Ramakrishnan, and A. Weixbaumer for comments on the original idea. K. Takai for sharing his knowledge about base modifications and wobble, and A. Buguin for suggestions of improvements. A fellowship from the Swiss National Science Foundation (Grant no 105059 to J.L.) and the Marie-Josée and Henry Kravis Fellowship at the Rockefeller University (to J.L.) are gratefully acknowledged.

Received February 8, 2008; accepted March 30, 2008.

**REFERENCES**


Degeneracy of the genetic code


Degeneracy of the genetic code and stability of the base pair at the second position of the anticodon

Jean Lehmann and Albert Libchaber

*RNA* 2008 14: 1264-1269