F1000Prime REPORTS

Published: 01 April 2014

Recognition of Watson-Crick base pairs: constraints and limits due to geometric selection and tautomerism

Eric Westhof[1]*, Marat Yusupov[2] and Gulnara Yusupova[2]

Addresses: [1]Architecture et Réactivité de l'ARN, Université de Strasbourg, Institut de Biologie Moléculaire et Cellulaire, CNRS, 15 rue René Descartes, F-67084 Strasbourg Cedex, France; [2]Département de Biologie et de Génomique Structurales, Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS, INSERM, Université de Strasbourg, F-67400 Illkirch, France

* Corresponding author: Eric Westhof (E.Westhof@ibmc-cnrs.unistra.fr)

F1000Prime Reports 2014, **6**:19 (doi:10.12703/P6-19)

Abstract

The natural bases of nucleic acids have a strong preference for one tautomer form, guaranteeing fidelity in their hydrogen bonding potential. However, base pairs observed in recent crystal structures of polymerases and ribosomes are best explained by an alternative base tautomer, leading to the formation of base pairs with Watson-Crick-like geometries. These observations set limits to geometric selection in molecular recognition of complementary Watson-Crick pairs for fidelity in replication and translation processes.

Introduction and context

Recognition between biopolymers (nucleic acids or proteins) requires tight and close packing in a specific fashion for both partners. Broadly speaking, while the association of molecules through van der Waals interactions promotes the close packing, the directed H-bonds formed by defined side chains on each partner guarantee the specificity of the complex. In the description of biomolecular recognition, intermolecular hydrogen bonding is generally emphasized because of its awesome precision and the ensuing straightforward conclusions concerning the effects of possible mutations or sequence variations. However, the contacts present in the assembled complex result from a delicate balance between lock-and-key accommodation and induced-fit mutual rearrangements in the components. The tightness of the final fit between the molecules is achieved through conformational changes in torsion angles distributed over several residues, some of them far away from the binding interface, and through inclusion or exclusion of water molecules and ions.

The recognition of Watson-Crick base pairs is at the core of the main molecular biology processes (replication, transcription, and translation). In a seminal paper, Seeman and collaborators [1] identified two main characteristics for the recognition of Watson-Crick pairs by another macromolecule: (a) two H-bonds are required for achieving fidelity of recognition; (b) recognition from the minor groove side is relatively insensitive to base pair reversals. In DNA replication, RNA transcription or ribosomal decoding, the recognition process must work equally well with any of the four possible Watson-Crick base pairs. Indeed, in all those cases, crystal structures of complexes show that recognition occurs in the minor groove of DNA (or the wide shallow groove of RNA) through the amino acid side chains in polymerases [2-4] and through ribosomal nucleotides in the ribosome [5].

DNA polymerase

DNA polymerases display a wide range in replication fidelity or error rate (number of wrong nucleotides inserted per event), with error rates higher than 10^{-2} in the Y-family polymerases and as low as around 10^{-9} (or one error in 1 billion inserted nucleotides) [6]. Such high fidelities in replication are achieved through three main processes: polymerase selectivity, proofreading and mismatch repair [7]. Proofreading-deficient polymerases still achieve a remarkable error rate for incorrect insertion between 10^{-3} and 10^{-6} depending on the type and environment [8]. It was soon realized that free energy differences between matched and mismatched base pairs (<4 kcal/mole) could not solely be responsible for the observed error rates [9],

which led to the concept of "geometric selection" [10,11]. The insight that "geometric selection" contributes several orders of magnitude to replication fidelity was brilliantly illustrated by the observation that the nonpolar but isosteric analog (like difluorotoluene, lacking good H-bonding capability, and is isosteric to thymine) could be very efficiently incorporated by a defective-proofreading DNA polymerase I [12]. The selection through geometry was elaborated in the concept of the active site tightness in which the selected shapes of isosteric Watson-Crick base pairs fit snugly into the active site of the enzyme and lead to enzymatic activity [13]. Several mutations increasing or decreasing fidelity could be rationalized by their effects on the grip of the Watson-Crick pair in the active site [13].

RNA polymerase

RNA polymerase is slightly less accurate, with error rates around 10^{-5} in bacteria and eukaryotes [14]. Like DNA polymerases, RNA polymerases exploit (after nucleotide selection) proofreading mechanisms, in which the RNA polymerase slows down after misincorporation with subsequent cleavage of the mismatched nucleotide [15-17]. Nucleotide selection involves an isomerization from an open to a closed state formed upon the folding of a trigger loop [15]. In yeast Pol III, which synthesizes noncoding and generally structured RNAs, both processes contribute significantly to an overall fidelity of 2.0×10^{-7}, with 1.8×10^{-4} brought by nucleotide selectivity alone before proofreading [18].

Ribosomal translation

Ribosomal translation leading to protein synthesis, although the least accurate of the fundamental processes, still presents the remarkable accuracy of only 1 error per 10^3–10^4 codons or amino acid inserted in bacteria [19]. The fidelity of translation depends strongly on the correct selection of the aminoacylated transfer RNA (tRNA) cognate to the codon being translated [20-22]. This is partly due to the fact that tRNA aminoacylation is a highly accurate process (1 error in 10^6 [23,24]). Again, aminoacyl-tRNA synthetases achieve such high accuracies through editing and proofreading mechanisms following catalysis [25].

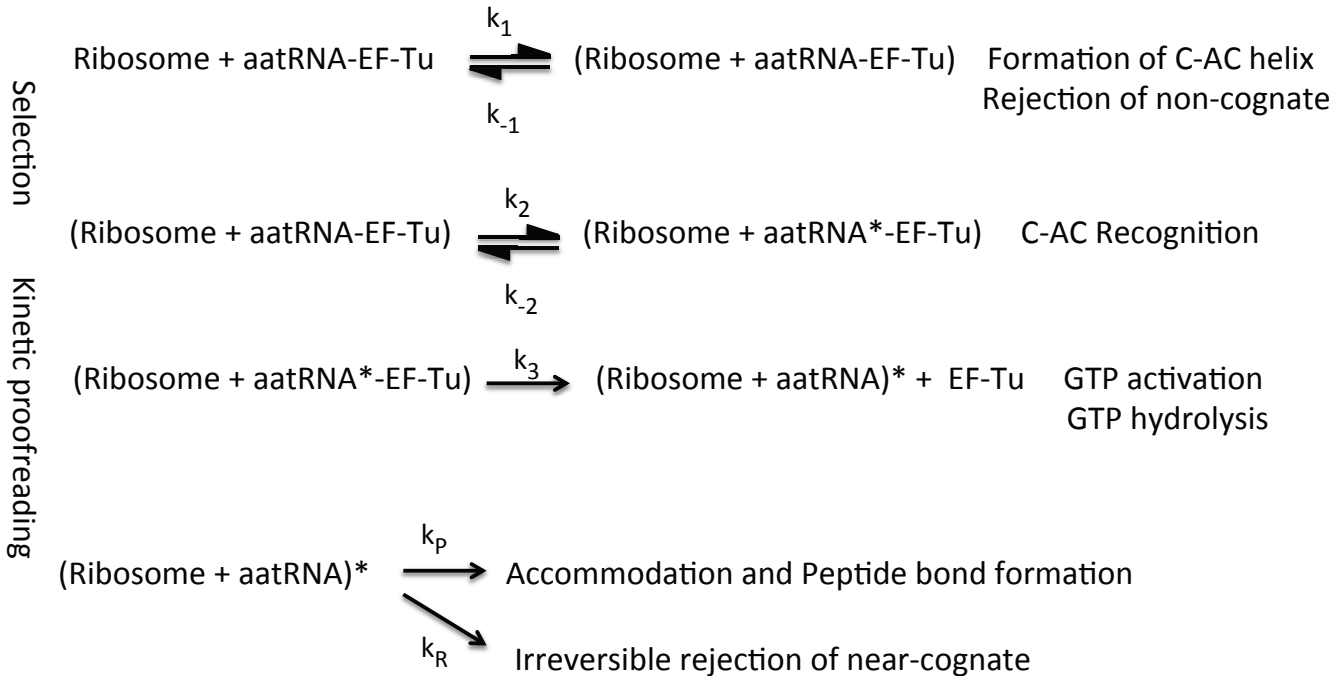
The discrimination steps for tRNA selection by the ribosome are spread over two main processes occurring in two different molecular environments: initial selection of the tRNA complexed to the elongation factor EF-Tu (elongation factor thermo unstable), and kinetic proofreading or amplification following GTP hydrolysis (for reviews see [22,26-29] and also see Figure 1). As shown by X-ray studies on the 30S ribosomal subunits, the nucleotides of the helix formed between the mRNA codon and the anticodon triplet of the cognate aminoacylated tRNA (the decoding helix) are contacted by several conserved residues of the small ribosomal subunit (especially G530, A1492, A1493) that all contact the minor groove side of the first two base pairs of the codon-anticodon helix by forming an A-minor motif [5,26]. The third base pair, or wobble position, is bound differently and asymmetrically by nucleotides G518, C530, and C1054. A-minor motifs monitor the minor groove side of RNA helices and bind preferentially complementary Watson-Crick base pairs [30].

For a given codon (in the A site of the ribosome), the probable outcome of an incoming aminoacylated tRNA depends on the number of non-standard Watson-Crick pairs (including any of the wobble pairs) in the first two base pairs of the codon-anticodon helix. This can be cognate (an expected tRNA with no non-standard Watson-Crick pairs, able to elicit GTP hydrolysis and peptide synthesis), near-cognate (an incorrect tRNA with generally a single non-standard Watson-Crick pair, still capable of eliciting GTP hydrolysis and, thus, potentially contributing to translation errors), or non-cognate (an incorrect tRNA with generally more than one non-standard Watson-Crick pair, leading to an unstable codon-anticodon helix, incapable of eliciting GTP hydrolysis and eliminated at the selection step). Several ribosomal mutations [21,31,32] or antibiotics like aminoglycosides [33] affect translation accuracy by increasing misreading of near-cognate tRNAs. Interestingly, recent work has demonstrated that, after a misincorporation, a quality control system induces a general loss of fidelity at the A site [34,35]. This post-peptidyl transfer control could be considered akin to editing cleavages in DNA polymerases or aminoacyl-tRNA synthetases.

Early discussions on decoding fidelity are dominated by geometric selection

It was soon realized that the free energy differences between complementary Watson-Crick pairs and the possible non-complementary pairs could not explain the observed accuracies, especially in the DNA polymerases [9]. For example, the lifetimes of cognate complexes were shown to be less than two orders of magnitude higher than the incorrect complexes [36] and affinities between cognate and non-cognate triplets differ by no more than a factor of ten [37]. Early theoretical work on the accuracy of ribosomal translation [38-40] and experimental data, especially on antibiotics affecting translation [41,42], emphasized geometric selection and steric constraints imposed by the ribosome in order to discriminate between complementary Watson-Crick pairs and non-complementary pairs. In parallel, an alternative, but not exclusive, mechanism was put forward by Hopfield [43] and Ninio [44] and soon after experimentally verified [45]: the kinetic proofreading or kinetic amplification in which a

Figure 1. Simplified schemes for the main kinetic recognition steps in ribosomal translation



Simplified kinetic schemes describing the transfer RNA (tRNA) discrimination process by the ribosome following current understanding [26,28,69,95]. The rate constant k_{-2} has been measured to be 1000-fold faster for near-cognate tRNAs than for cognate tRNAs (corresponding to a $\Delta\Delta G$ around 4.2 kcal/mole), while the rate constant k_3 is 500-fold faster for cognate than for near-cognate tRNAs [69]. Note that induced fits and conformational selection are difficult to distinguish kinetically, and that both imply, according to classical enzymology, the inclusion of a four-membered cycle within the reaction scheme [96], a feature that has not been implemented into the present schemes.

proofreading step occurs after initial selective binding consumes GTP.

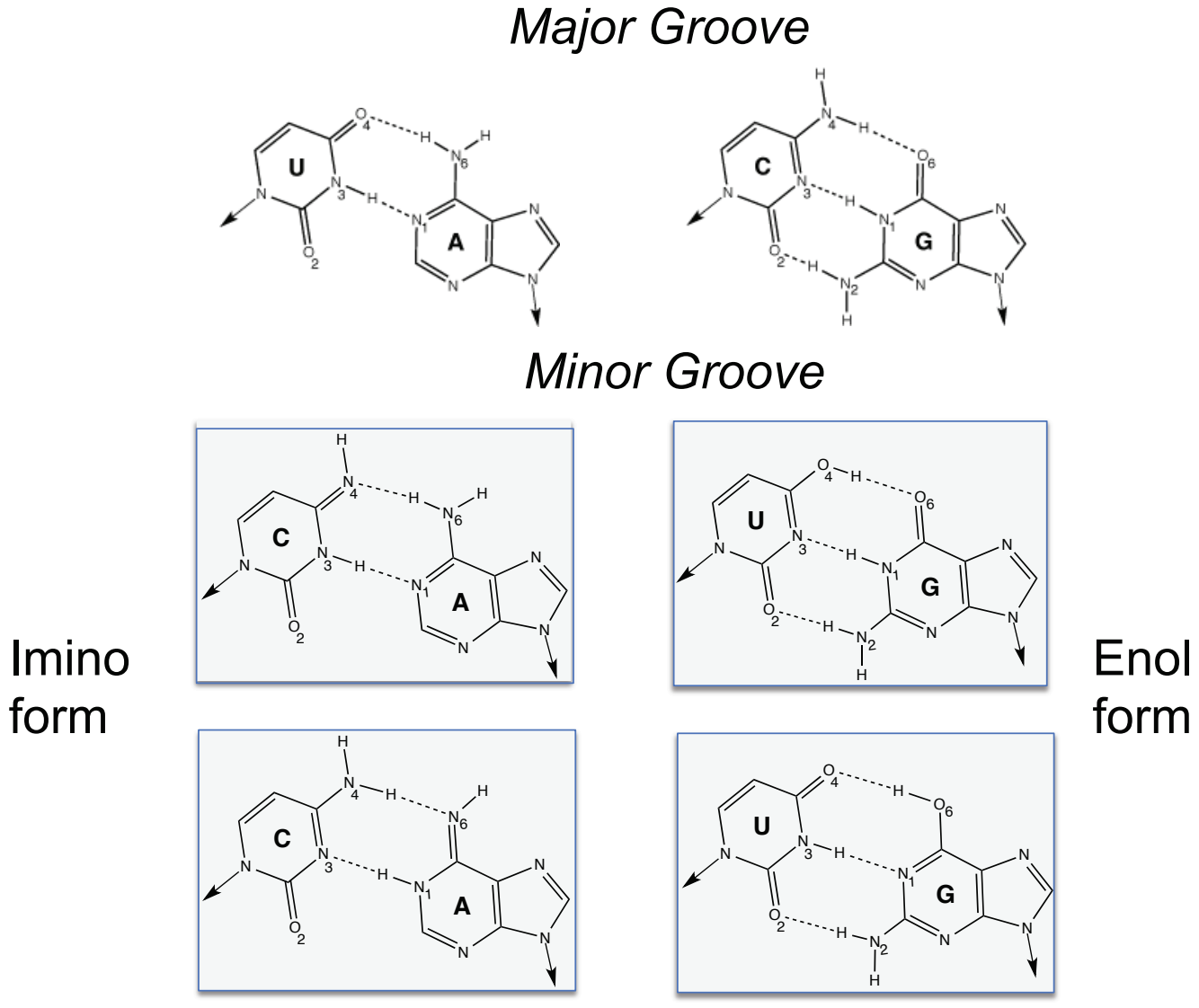
The four natural nucleic acid bases (A, G, C, U) are characterized by their highly preferred tautomeric form, so central to precise and regular recognition, with their minor tautomers present only in ratios around 1 for 10^3 or 10^4 standard states. The simple position exchange of the amino and keto groups in G or in C (giving isoguanine [iso-G] or isocytosine [iso-C], respectively) yields bases with highly ambivalent tautomerism, iso-G, or with too facile deamination reactions, iso-C [46]. This observation led to the conclusion that the iso-G-iso-C base pair could not have been an information storage molecule in early molecular evolution billion of years ago especially when in competition with G=C and A-U pairs [46]. In 1976, Topal and Fresco published two groundbreaking articles on base pairing recognition in replication [47] and in translation [39]. They widened the concept of complementarity and analyzed with great insight the consequences of base tautomerism in both processes. In 1953, Watson and Crick [48] had already proposed that spontaneous mutations might occur when pairs are formed with one base in a rare tautomeric form. Indeed, with a keto-enol tautomerism on either base, both the C∼A/A∼C pairs and U∼G/G∼U pairs display exactly the same dimensions as the standard complementary pairs C=G/G=C or U-A/A-U (Figure 2), unlike the wobble pairs UoG/GoU in which the pyrimidine is displaced in the major groove, creating a small cavity on the minor groove side.

In their second article [39], Topal and Fresco discuss the base pairing schemes, some of which involve tautomerism, that possess the dimensions and shapes close to the complementary Watson-Crick pairs so that they can be accommodated by or pass through the sieve formed by the steric and geometric constraints imposed by the ribosome. They stress the point that, while formation of unfavored tautomers would not occur once the nucleotides are within the ribosomal cavity (mainly because of water exclusion), unfavored tautomers formed, before being closed up of the ribosomal cavity and, according to solution equilibria, would be locked in.

Recent advances in replication and decoding ferret out tautomerism

Two recent articles give strong structural support to the role of tautomerism in replication infidelity with the

Figure 2. Standard Watson-Crick pairs and related tautomeric pairs



Standard complementary U-A (left) and C=G (right) Watson-Crick pairs and the isosteric C~A and U~G pairs formed using the tautomers imino (for C and A – left) and enol (for U and G – right). The symbol ~ has been used throughout for noting a pair involving a tautomer. The environments of the major (deep) and minor (shallow) grooves of the standard Watson-Crick pairs are the same in both types of the tautomeric forms (imino and enol) for the C~A and U~G pairs.

observation of a G~T base pair and a C~A base pair in a Watson-Crick-like geometry trapped in an active state of a polymerase, respectively in a human DNA polymerase λ variant and in the *Bacillus stearothermophilus* DNA polymerase I large fragment [4,49]. In order to be able to trap crystallographically, these rare mismatched events, a polymerase mutant with a five amino acid deletion (that does not impair catalysis or correct nucleotide insertion) was used for crystallization in the first case and, in the second case, Mg^{2+} ions were substituted by the mutagenic Mn^{2+} ions.

Recently, unexpected base pairing states have been observed in crystal structures of 70S bacterial ribosomes primed with a 30 nt mRNA and with the A-, P- and E-sites occupied by tRNAs where the A-site tRNAs were near-cognate: in each crystal structure one of the three pairs of the codon-anticodon triplet was not a standard Watson-Crick pair [50,51]. Indeed, with near-cognate tRNAs presenting a G in front of a U at either the first or second triplet position, one observes a Watson-Crick-like geometry of the U~G/G~U pairs, which can best be rationalized by the formation of a keto-enol tautomer of either base. On the

other hand, with a tRNA where the G in front of the U is at the third, or wobble, position of the codon-anticodon triplet, the U and G residues adopt the expected GoU wobble pair [52] with the U pushed into the major groove. The observed final states of the ribosomes with near-cognate or cognate tRNAs were similar, implying that the ribosomes containing near-cognate tRNAs were in an active conformation poised for GTP hydrolysis followed by accommodation and peptide bond formation.

These new observations [50,51], together with the recent ones on the DNA polymerases [4,49], emphasize that these complex molecular machineries recognize, first of all, the shapes of the base pairs and not the numbers and types of hydrogen bonds that form them. In this respect, recognition rules by these enzymes are different from those governing stabilities of double-stranded nucleic acids with natural or non-natural bases [53].

Base pairs with isosteric shapes to the usual complementary Watson-Crick pairs can be obtained in three main ways: by tautomerism in one of the bases [39], by a non-natural non-polar residue complementary to a standard base [12,13,54], or by a mixture of both in some nucleotide analogs that can pair to standard bases [55]. In addition, the recognition of the base pair shapes, in polymerases and ribosomes, occurs mainly around the minor groove side and, since the Watson-Crick-like pairs involving tautomerism, U∼G/G∼U and C∼A/A∼C, offer in the minor groove the same disposition of H-bond donors and acceptors, the binding interface is preserved, further fooling the recognition process (see Figure 2).

Crystallogenesis conforms to thermodynamics

In another realm of the RNA world, the interplay between extreme accuracy in binding together and the remarkable potential for adaptability can be observed. Riboswitches are segments of noncoding RNAs generally found in the five prime untranslated region (5'UTR) of genes they control. In the presence of a defined ligand, the specific riboswitch will adopt an intricate RNA architecture that encapsulates the ligand with high discriminatory power [56]. Recent structural and thermodynamic works (for example [57] and discussion in [58]) demonstrate beautifully how, despite the surrounding tightness of the binding site, slightly modified ligand analogs can still be recognized and bound at small free energy costs with minor tautomer changes in the ligands.

All the structural data discussed here are based on the X-ray crystallography of crystals of these highly complex machineries. Crystallographers make huge efforts to obtain crystallized complexes as biologically relevant as possible and to produce the necessary biological controls. So here, one could rightly wonder how it is possible to trap, long enough to generate a crystal, an event occurring once in 10,000 events. Although crystallography is not thermodynamics, crystallogenesis has to follow thermodynamics. One approach to trap unfavorable or rare events consists of mutating key elements or adding different coions or antibiotics, all of which were previously observed to decrease the fidelity of the process. All of these factors have been used to various degrees in the experiments discussed here. Two other factors should be considered. Firstly, the crystallization conditions are such that there is no competition between correct and incorrect ligands (consequently, by the law of mass action, unfavorable free energy differences can be overcome). Secondly, not unrelated to the preceding factor, the crystallization conditions have been painfully established often over many years so that the system is driven to reproducible and diffracting crystals. Again, some loss in the free energy of binding can be compensated for by other free energy components in the overall free energy. In crystals of riboswitches, it is regularly observed that the unliganded architectures are very close to the bound ones [59,60]. In short, rare tautomeric base pairs will be favored if their formation leads to a reduction in the overall free energy of the macromolecular system. Complexes with near-cognate ligands reveal how the ligands are accommodated in the active site. In other words, they are views of states after ligand selection is completed and, only indirectly, do they provide information about how discrimination between correct and incorrect ligands is achieved.

Further consequences and debates

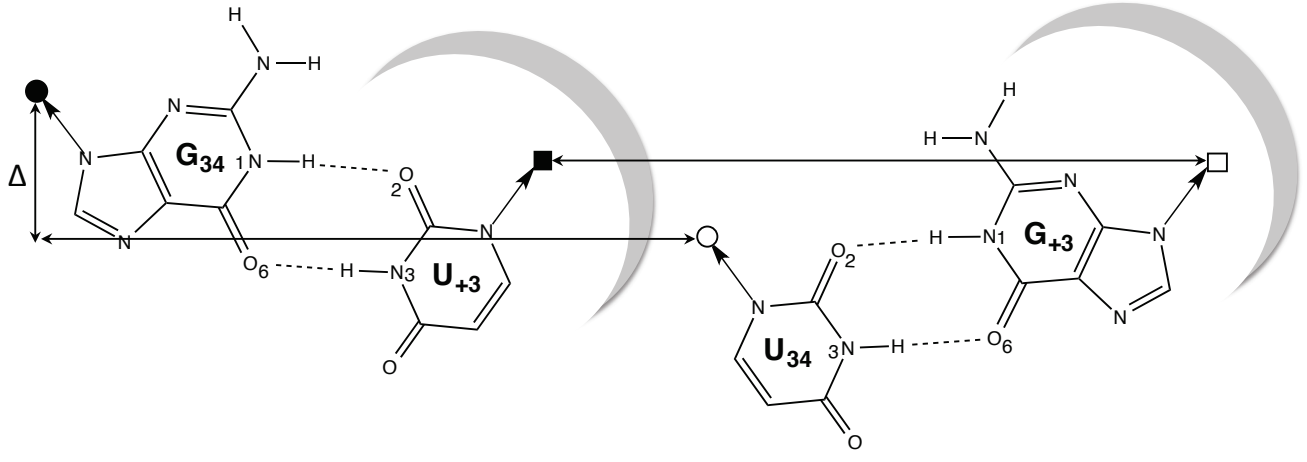
Among the three base pairs formed in the codon-anticodon helix, the third one, or wobble pair between the third mRNA base and the first nucleotide (residue 34) of the anticodon triplet [52], is very special for numerous reasons related to the degeneracy of the codon table. Nucleotide 34, advantageously located at the apical tip of the anticodon hairpin, has some "wobbling" capacity leading to possible non-standard pairings. Because of the relaxed constraints on the third base pair, much less than 61 different tRNAs are needed to decode the 61 sense codons. However, in tRNAs, nucleotide 34 is the most frequently modified residue with a great diversity of chemical modifications (adenosine is changed into inosine, U is modified at position C5 and/or O2, G at position N7) throughout the phylogenetic tree [61]. With regard to that base pair, two points will be discussed here: firstly, how the ribosome deals with the non-isostericity of GoU pairs upon reversal; secondly, how the base pairs involving inosine (AoI and GoI) are accommodated. Both points implicate geometric selection and base tautomerism. Structurally, UoG34 pairs are observed as standard wobble pairs [5,26,39,52,62,63]. Interestingly, recent results [64,65] confirm earlier findings

[66], showing that codons with standard Watson-Crick pairs at all three positions are translated faster than those with a wobble pair at the third position. This is also reflected in the 2.5 times higher rate of hydrolysis of GTP for fully Watson-Crick paired codons, compared to cognate tRNAs with a wobble pair at the third position [67-69].

An underappreciated property of wobble pairs is that they are not isosteric upon GoU to UoG reversal. This is particularly relevant if one of the two nucleotides is constrained in an active site, as is the case for nucleotide +3 of the mRNA, which is fixed in the A site of the ribosome to ribosomal nucleotides and protein S12 via a magnesium ion (Figure 3). Thus, for accommodation of any non-standard pair, movements can essentially occur at tRNA residue 34. Because of the tRNA anticodon loop fold, movements towards the minor groove (necessary for a G34oU[+3] pair) are easier than those towards the major groove (required for a U34oG[+3] pair). tRNAs manage, however, by several modifications of residue U34 that stabilize a tautomer change, to form a U34~G(+3) with Watson-Crick-like geometry [70-73]. Reversibly, some U34 modifications, although promoting U34~G(+3) formation, are detrimental to the formation of the standard Watson-Crick U34-A(+3) pair [74]. The importance of U34 modifications for reading G-ending codons has been emphasized several times [75,76]. This is not the only known case where a modification promotes a tautomer form leading to a Watson-Crick-like geometry pair: both 2-agmatinylcytidine or 2-lysylcytidine (lysidine) exist in a tautomer form, guaranteeing pairing with A, so that the isoleucine AUA codon is translated instead of the Methionine AUG codon in bacteria (which is translated with the unmodified CAU anticodon) [77].

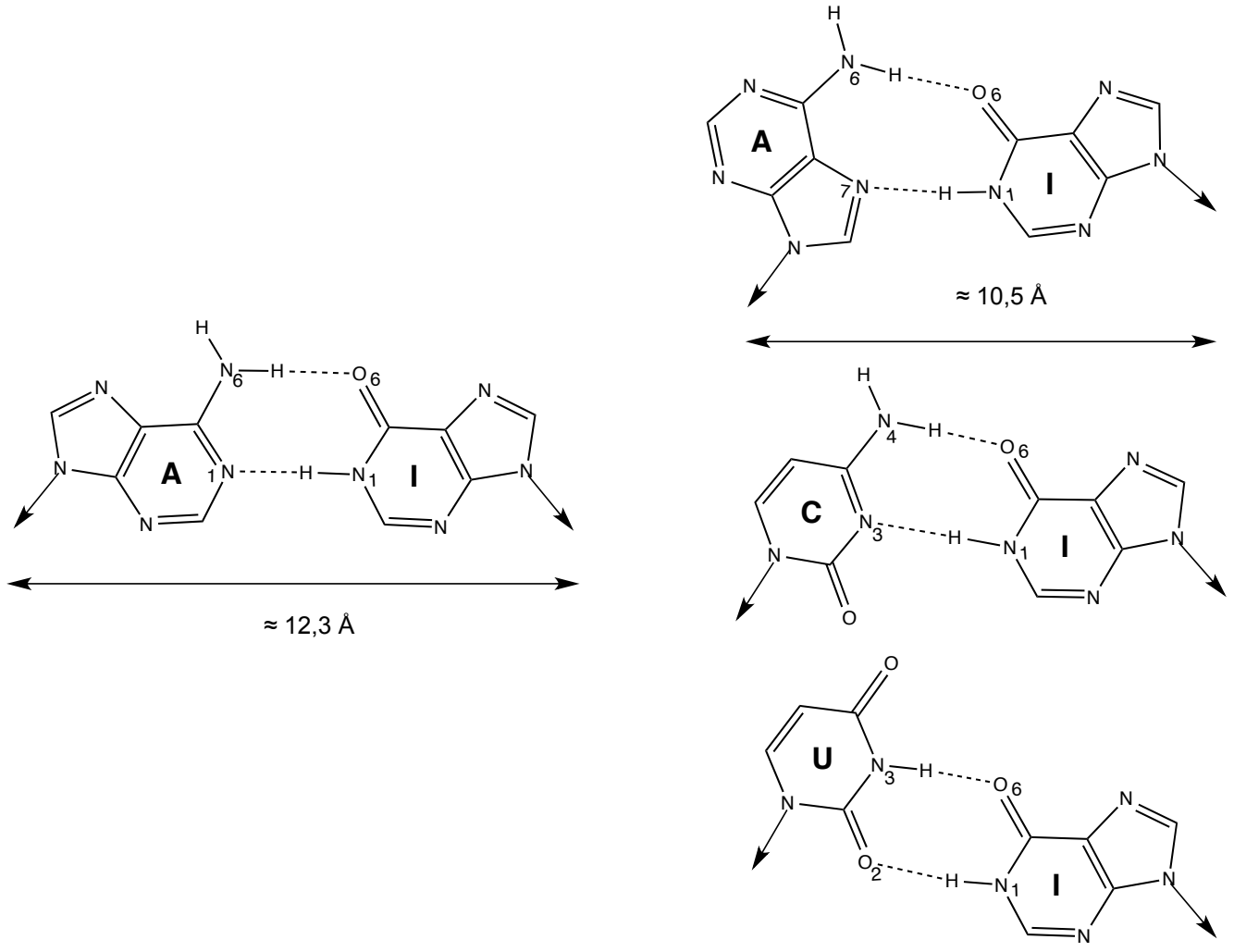
A last base-pair type occurring at the third position, source of multiple controversies [78], is worth discussing: the IoA and IoG base pairs formed by inosine (that replaces adenosine at position 34). These pairs are purine-purine opposition with a distance between the ribose C1′ atoms around 12.3 Å instead of the usual 10.5 Å typical of standard Watson-Crick pairs (Figure 4). The observations, structural and kinetic, described and reviewed above converge on the central role of the ribosomal grip at the decoding center, moulding and constraining Watson-Crick base pairing for all three base pairs of the anticodon-codon triplet helix, and thus a C1′-C1′ distance around 10.5 Å. In 1976, a similar conclusion was reached by Topal and Fresco [39] who concluded that a Watson-Crick/Hoogsteen pair (for nomenclature, see [79]) between A and I was "more reasonable". Such a pair has the right C1′-C1′ distance but requires that the A be in the *syn* conformation of the base with respect to the ribose. Although previous work [78] had decided on the presence of the long purine-purine pair, a very recent crystal structure [80] contained such a Watson-Crick/Hoogsteen pair between a G and A*syn*, but at the second position. That same crystal structure displayed (at the third position) an I residue pairing through the Watson-Crick side with the Hoogsteen face of a G*syn* base. It has been proposed that a magnesium ion stabilizes the IoG*syn* base pair. However, an isosteric base pair to IoG*syn* could as well form under other

Figure 3. The non-isostericity of the GoU wobble pairs



The wobble base pairs, denoted GoU and UoG, are not isosteric upon reversal. Each third residue of the mRNA is locked by ribosome contacts (indicated by the greyish crescent) and thus only the wobble base 34 in the tRNA anticodon has some mobility. However, from a G34 to a U34, there is a large distance (around 4.0 Å) that the tRNA loop cannot well achieve. The formation of a Watson-Crick configuration is thus favored and this is often accomplished by specific modifications of the U34.

Figure 4. Some pairs involving inosine



Base pairs involving the modified nucleotide inosine, found instead of A at position 34 in tRNAs. The distance between the ribose C1' atoms in the Watson-Crick/Watson-Crick A=I pair (with both bases in the anti conformation with respect to the ribose) is around 12.3 Å (left), much larger than the usual 10.5 Å. However, with the Hoogsteen of A H-bonding to the Watson-Crick edge of I, a pair can be formed with the standard distance (right). In that case, because of the helical conformation of the triplet helix, the A should in the *syn* conformation. Note that UoI forms a wobble type of pair.

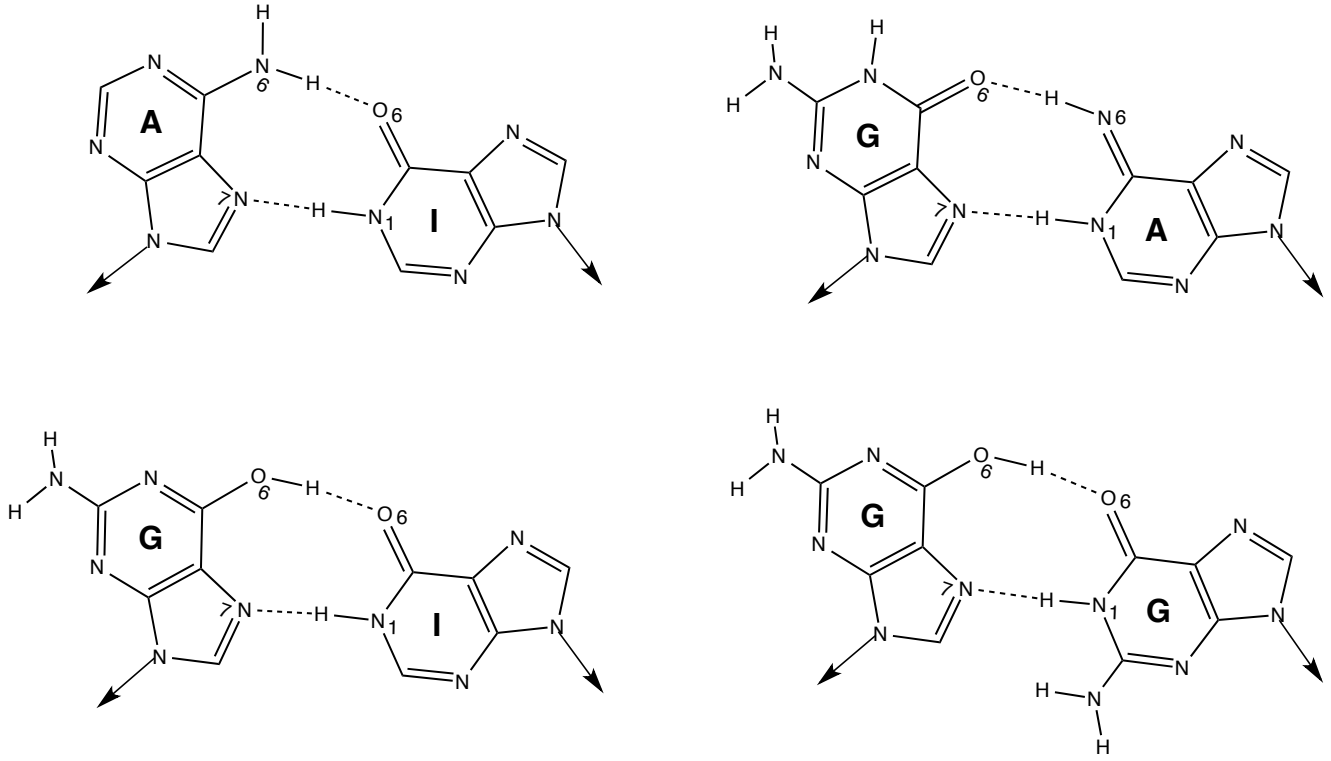
conditions with a G in the *enol* tautomer, as proposed by Topal and Fresco [39] (Figure 5).

The observed accommodation of unusual base pairs within the decoding center is less due to an "unsuspected plasticity" [80] of the ribosome than to the preference of the decoding center for the usual shape and geometry of Watson-Crick base pairs. However, the numerous and diverse molecular interactions present within nucleotides (conformational equilibria like *anti/syn*, tautomerism, chemical modifications) allow for enough molecular adaptability for the formation of sets of isosteric base pairs. Thus, even with the tight size and shape controls exerted by the ribosome, unusual base pairs can be trapped giving rise to alternative decoding [77,81] or translational errors.

Future directions

Discussion (and controversies) concerning the relative roles of hydrogen bonding even in fluorine-containing analogs are still continuing to enrich our scientific understanding [54,82]. It would be most illuminating to see a crystal structure with a difluorotoluene trapped in an active state of a polymerase. By extension, the use of difluorotoluene in studies of ribosomal translation might be fundamentally and practically useful. Experiments in this direction, using 2'fluoro or 2deoxy mRNAs, have been recently reported [83]. They show that 2'fluoro

Figure 5. Some isosteric purine-purine pairs



A set of four isosteric purine-purine base pairs of the family *cis* Watson-Crick/Hoogsteen. For each of them, the base at the left should be in the *syn* conformation and the base at the right in the usual *anti* conformation with respect to the ribose. Three of them involve a tautomer of one of the two bases. (Left) A and G with inosine (similar drawings can be made replacing I by G). For translational decoding, I should be residue 34 of the anticodon and A/G at the third position on the message. (Right) G (on the message) with either A or G (on the anticodon).

mRNAs have a 1000-fold difference with 2′deoxy mRNAs in the rate of peptide bond formation (and that the triple 2′deoxy mRNAs are largely rescued by the aminoglycoside paramomycin). These observations led the authors to the inescapable conclusion that steric complementarity and shape recognition are more important in the decoding center than hydrogen bonding [50,83,84].

A limitation to further work and understanding is the lack of definite experimental knowledge of the tautomeric equilibria of the natural bases, although many theoretical calculations (often in the gas phase) are available [85]. Free energy differences between the keto/amino forms of the bases and their tautomers are reported in the range between 5 and 7 kcal/mole, corresponding to frequencies of occurrence of 1 in 10^4 or 10^5 [39], and such values are not far from the accepted error rate in translation of 1 in 10^3 to 10^4. It has already been noted by Topal and Fresco [39] that "*in vivo* infidelity is of a level that might be expected from the frequency of the minor tautomers in both message and tRNA before their interaction within the ribosome", leading them to conclude that "there is no obvious basis and no apparent need for proofreading of infidelity caused by complementary mispairs". However, intrinsic selectivities (as deduced from free energy differences between correct and incorrect ligands) constitute upper limits to accuracies [86] and the use of GTP analogs that slow down nucleotide hydrolysis led to the conclusion that proofreading is not only there for accuracy but also to maintain the most adequate tradeoff between efficiency and accuracy [20,67,87,88]. Relative cellular tRNA concentrations in the overall cellular environment play further additional roles in the strongly interconnected processes underlying ribosomal protein synthesis [21,22].

Errors open doors for evolution and tinkering

Precision in molecular architectures and specific binding is anchored in various defined physico-chemical atomic interactions. Because of the intrinsic neutrality of those molecular interactions (still heavily constrained by the selected nucleic acid bases and protein constituents), biomolecular architectures can accommodate errors, thereby participating in their own evolution as well as in the

construction of highly adaptable and robust biochemical networks.

One can hope that the present considerations will contribute also to the *in vitro* adaptation of bacterial ribosomal synthesis of unnatural peptides [89] and peptidomimetics [90], so useful nowadays for novel reagents in biotechnology and drug discovery in therapy. All these innovative techniques manipulate ribosomal processes, especially decoding [91-94], and require a clear understanding of the range of errors that can slip in through the various fidelity control mechanisms, their origins and accommodations.

Abbreviations

5′ UTR, five prime untranslated region; Iso-C, isocytosine; iso-G, isoguanine; tRNA, transfer RNA.

Disclosures

The authors declare that they have no disclosures.

Acknowledgments

The literature on ribosomal fidelity is extremely vast and spans several decades of dedicated work in many laboratories around the world by many researchers. The aim of this review was to address the role of steric recognition and tautomerism in structural biology and was not intended as an exhaustive review of all observations and literature on ribosomal translation and errors. We apologize if we have not included all relevant publications. The authors are grateful to Valérie Fritsch for the molecular drawings and to their colleagues for a very fruitful collaborative thinking and discussions on ribosomal decoding. Eric Westhof would like to thank Jacques Ninio for numerous enlightening exchanges on the kinetic aspects of recognition fidelity. This work has been published under the framework of the LABEX: ANR-10-LABX-0036_NETRNA and benefits from a funding from the state managed by the French National Research Agency as part of the program "Investments for the future".

References

1. Seeman NC, Rosenberg JM, Rich A: **Sequence-specific recognition of double helical nucleic acids by proteins.** *Proc Natl Acad Sci USA* 1976, **73:**804-8.

F1000Prime RECOMMENDED

2. Doublié S, Tabor S, Long AM, Richardson CC, Ellenberger T: **Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution.** *Nature* 1998, **391:**251-8.
3. Evans RJ, Davies DR, Bullard JM, Christensen J, Green LS, Guiles JW, Pata JD, Ribble WK, Janjic N, Jarvis TC: **Structure of PolC reveals unique DNA binding and fidelity determinants.** *Proc Natl Acad Sci USA* 2008, **105:**20695-700.

F1000Prime RECOMMENDED

4. Wang W, Hellinga HW, Beese LS: **Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis.** *Proc Natl Acad Sci USA* 2011, **108:**17644-8.

F1000Prime RECOMMENDED

5. Ogle JM, Brodersen DE, Clemons WM, Tarry MJ, Carter AP, Ramakrishnan V: **Recognition of cognate transfer RNA by the 30S ribosomal subunit.** *Science* 2001, **292:**897-902.

F1000Prime RECOMMENDED

6. McCulloch SD, Kunkel TA: **The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases.** *Cell Res* 2008, **18:**148-61.
7. Kunkel TA: **Evolving views of DNA replication (in)fidelity.** *Cold Spring Harb Symp Quant Biol* 2009, **74:**91-101.
8. Kunkel TA, Bebenek K: **DNA replication fidelity.** *Annu Rev Biochem* 2000, **69:**497-529.
9. Loeb LA, Kunkel TA: **Fidelity of DNA synthesis.** *Annu Rev Biochem* 1982, **51:**429-57.
10. Echols H, Goodman MF: **Fidelity mechanisms in DNA replication.** *Annu Rev Biochem* 1991, **60:**477-511.
11. Goodman MF: **Hydrogen bonding revisited: geometric selection as a principal determinant of DNA replication fidelity.** *Proc Natl Acad Sci USA* 1997, **94:**10493-5.
12. Moran S, Ren RX, Kool ET: **A thymidine triphosphate shape analog lacking Watson-Crick pairing ability is replicated with high sequence selectivity.** *Proc Natl Acad Sci USA* 1997, **94:**10506-11.
13. Kool ET: **Active site tightness and substrate fit in DNA replication.** *Annu Rev Biochem* 2002, **71:**191-219.

F1000Prime RECOMMENDED

14. Sydow JF, Cramer P: **RNA polymerase fidelity and transcriptional proofreading.** *Curr Opin Struct Biol* 2009, **19:**732-9.
15. Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD: **Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis.** *Cell* 2006, **127:**941-54.

F1000Prime RECOMMENDED

16. Sydow JF, Brueckner F, Cheung ACM, Damsma GE, Dengl S, Lehmann E, Vassylyev D, Cramer P: **Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA.** *Mol Cell* 2009, **34:**710-21.

F1000Prime RECOMMENDED

17. Vassylyev DG, Vassylyeva MN, Zhang J, Palangat M, Artsimovitch I, Landick R: **Structural basis for substrate loading in bacterial RNA polymerase.** *Nature* 2007, **448:**163-8.
18. Alic N, Ayoub N, Landrieux E, Favry E, Baudouin-Cornu P, Riva M, Carles C: **Selectivity and proofreading both contribute significantly to the fidelity of RNA polymerase III transcription.** *Proc Natl Acad Sci USA* 2007, **104:**10400-5.

F1000Prime RECOMMENDED

19. Parker J: **Errors and alternatives in reading the universal genetic code.** *Microbiol Rev* 1989, **53:**273-98.
20. Wohlgemuth I, Pohl C, Mittelstaet J, Konevega AL, Rodnina MV: **Evolutionary optimization of speed and accuracy of decoding on the ribosome.** *Philos Trans R Soc Lond, B, Biol Sci* 2011, **366:**2979-86.
21. Kramer EB, Farabaugh PJ: **The frequency of translational misreading errors in E. coli is largely determined by tRNA competition.** *RNA* 2007, **13:**87-96.

F1000Prime RECOMMENDED

22. Zaher HS, Green R: **Fidelity at the molecular level: lessons from protein synthesis.** *Cell* 2009, **136:**746-62.

23. Söll D: **The accuracy of aminoacylation–ensuring the fidelity of the genetic code.** *Experientia* 1990, **46:**1089-96.

F1000Prime RECOMMENDED

24. Francklyn CS: **DNA polymerases and aminoacyl-tRNA synthetases: shared mechanisms for ensuring the fidelity of gene expression.** *Biochemistry* 2008, **47:**11695-703.

25. Sankaranarayanan R, Moras D: **The fidelity of the translation of the genetic code.** *Acta Biochim Pol* 2001, **48:**323-35.

26. Ogle JM, Ramakrishnan V: **Structural insights into translational fidelity.** *Annu Rev Biochem* 2005, **74:**129-77.

27. Ninio J: **Multiple stages in codon-anticodon recognition: double-trigger mechanisms and geometric constraints.** *Biochimie* 2006, **88:**963-92.

28. Daviter T, Gromadski KB, Rodnina MV: **The ribosome's response to codon-anticodon mismatches.** *Biochimie* 2006, **88:**1001-11.

29. Zaher HS, Green R: **Hyperaccurate and error-prone ribosomes exploit distinct mechanisms during tRNA selection.** *Mol Cell* 2010, **39:**110-20.

30. Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA: **RNA tertiary interactions in the large ribosomal subunit: the A-minor motif.** *Proc Natl Acad Sci USA* 2001, **98:**4899-903.

F1000Prime RECOMMENDED

31. Triman KL: **Mutational analysis of the ribosome.** *Adv Genet* 2007, **58:**89-119.

32. Fagan CE, Dunkle JA, Maehigashi T, Dang MN, Devaraj A, Miles SJ, Qin D, Fredrick K, Dunham CM: **Reorganization of an intersubunit bridge induced by disparate 16S ribosomal ambiguity mutations mimics an EF-Tu-bound state.** *Proc Natl Acad Sci USA* 2013, **110:**9716-21.

33. Pape T, Wintermeyer W, Rodnina MV: **Conformational switch in the decoding region of 16S rRNA during aminoacyl-tRNA selection on the ribosome.** *Nat Struct Biol* 2000, **7:**104-7.

34. Zaher HS, Green R: **Quality control by the ribosome following peptide bond formation.** *Nature* 2009, **457:**161-6.

F1000Prime RECOMMENDED

35. Zaher HS, Green R: **Kinetic basis for global loss of fidelity arising from mismatches in the P-site codon:anticodon helix.** *RNA* 2010, **16:**1980-9.

F1000Prime RECOMMENDED

36. Grosjean HJ, de Henau S, Crothers DM: **On the physical basis for ambiguity in genetic coding interactions.** *Proc Natl Acad Sci USA* 1978, **75:**610-4.

37. Kurland CG, Rigler R, Ehrenberg M, Blomberg C: **Allosteric mechanism for codon-dependent tRNA selection on ribosomes.** *Proc Natl Acad Sci USA* 1975, **72:**4248-51.

38. Eigen M, de Maeyer L: **Chemical means of information storage and readout in biological systems.** *Naturwissenschaften* 1966, **53:**50-7.

39. Topal MD, Fresco JR: **Base pairing and fidelity in codon-anticodon interaction.** *Nature* 1976, **263:**289-93.

F1000Prime RECOMMENDED

40. Potapov AP: **A stereospecific mechanism for the aminoacyl-tRNA selection at the ribosome.** *FEBS Lett* 1982, **146:**5-8.

41. Davies J, Gilbert W, Gorini L: **Streptomycin, suppression, and the code.** *Proc Natl Acad Sci USA* 1964, **51:**883-90.

42. Hornig H, Woolley P, Lührmann R: **Decoding at the ribosomal A site: antibiotics, misreading and energy of aminoacyl-tRNA binding.** *Biochimie* 1987, **69:**803-13.

43. Hopfield JJ: **Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity.** *Proc Natl Acad Sci USA* 1974, **71:**4135-9.

F1000Prime RECOMMENDED

44. Ninio J: **Kinetic amplification of enzyme discrimination.** *Biochimie* 1975, **57:**587-95.

F1000Prime RECOMMENDED

45. Thompson RC, Stone PJ: **Proofreading of the codon-anticodon interaction on ribosomes.** *Proc Natl Acad Sci USA* 1977, **74:**198-202.

F1000Prime RECOMMENDED

46. Switzer CY, Moroney SE, Benner SA: **Enzymatic recognition of the base pair between isocytidine and isoguanosine.** *Biochemistry* 1993, **32:**10489-96.

F1000Prime RECOMMENDED

47. Topal MD, Fresco JR: **Complementary base pairing and the origin of substitution mutations.** *Nature* 1976, **263:**285-9.

F1000Prime RECOMMENDED

48. Watson JD, Crick FH: **Genetical implications of the structure of deoxyribonucleic acid.** *Nature* 1953, **171:**964-7.

49. Bebenek K, Pedersen LC, Kunkel TA: **Replication infidelity via a mismatch with Watson-Crick geometry.** *Proc Natl Acad Sci USA* 2011, **108:**1862-7.

F1000Prime RECOMMENDED

50. Demeshkina N, Jenner L, Westhof E, Yusupov M, Yusupova G: **A new understanding of the decoding principle on the ribosome.** *Nature* 2012, **484:**256-9.

F1000Prime RECOMMENDED

51. Demeshkina N, Jenner L, Westhof E, Yusupov M, Yusupova G: **New structural insights into the decoding mechanism: translation infidelity via a G·U pair with Watson-Crick geometry.** *FEBS Lett* 2013, **587:**1848-57.

52. Crick FH: **Codon–anticodon pairing: the wobble hypothesis.** *J Mol Biol* 1966, **19:**548-55.

53. Geyer CR, Battersby TR, Benner SA: **Nucleobase pairing in expanded Watson-Crick-like genetic information systems.** *Structure* 2003, **11:**1485-98.

54. Kool ET, Sintim HO: **The difluorotoluene debate–a decade later.** *Chem Commun (Camb)* 2006:3665-75.

55. Harris VH, Smith CL, Jonathan Cummins W, Hamilton AL, Adams H, Dickman M, Hornby DP, Williams DM: **The effect of tautomeric constant on the specificity of nucleotide incorporation during DNA replication: support for the rare tautomer hypothesis of substitution mutagenesis.** *J Mol Biol* 2003, **326:**1389-401.

56. Breaker RR: **Riboswitches and the RNA world.** *Cold Spring Harb Perspect Biol* 2012, **4**.

57. Gilbert SD, Reyes FE, Edwards AL, Batey RT: **Adaptive ligand binding by the purine riboswitch in the recognition of guanine and adenine analogs.** *Structure* 2009, **17:**857-68.

F1000Prime RECOMMENDED

58. Fritsch V, Westhof E: **Molecular adaptation in RNA complexes.** *Structure* 2009, **17:**784-6.

59. Stoddard CD, Montange RK, Hennelly SP, Rambo RP, Sanbonmatsu KY, Batey RT: **Free state conformational sampling of the SAM-I riboswitch aptamer domain.** *Structure* 2010, **18:**787-97.

60. Wedekind JE: **The apo riboswitch as a molecular hydra.** *Structure* 2010, **18:**757-8.

61. Grosjean H, de Crécy-Lagard V, Marck C: **Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes.** *FEBS Lett* 2010, **584:**252-64.

62. Jenner L, Demeshkina N, Yusupova G, Yusupov M: **Structural rearrangements of the ribosome at the tRNA proofreading step.** *Nat Struct Mol Biol* 2010, **17:**1072-8.

63. Jenner LB, Demeshkina N, Yusupova G, Yusupov M: **Structural aspects of messenger RNA reading frame maintenance by the ribosome.** *Nat Struct Mol Biol* 2010, **17:**555-60.

64. Stadler M, Fire A: **Wobble base-pairing slows in vivo translation elongation in metazoans.** *RNA* 2011, **17:**2063-73.

F1000Prime RECOMMENDED

65. Spencer PS, Siller E, Anderson JF, Barral JM: **Silent substitutions predictably alter translation elongation rates and protein folding efficiencies.** *J Mol Biol* 2012, **422:**328-35.

F1000Prime RECOMMENDED

66. Sørensen MA, Pedersen S: **Absolute in vivo translation rates of individual codons in Escherichia coli. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate.** *J Mol Biol* 1991, **222:**265-80.

67. Thompson RC, Karim AM: **The accuracy of protein biosynthesis is limited by its speed: high fidelity selection by ribosomes of aminoacyl-tRNA ternary complexes containing GTP[gamma S].** *Proc Natl Acad Sci USA* 1982, **79:**4922-6.

68. Thomas LK, Dix DB, Thompson RC: **Codon choice and gene expression: synonymous codons differ in their ability to direct aminoacylated-transfer RNA binding to ribosomes in vitro.** *Proc Natl Acad Sci USA* 1988, **85:**4242-6.

69. Gromadski KB, Daviter T, Rodnina MV: **A uniform response to mismatches in codon-anticodon complexes ensures ribosomal fidelity.** *Mol Cell* 2006, **21:**369-77.

70. Murphy FV, Ramakrishnan V, Malkiewicz A, Agris PF: **The role of modifications in codon discrimination by tRNA(Lys)UUU.** *Nat Struct Mol Biol* 2004, **11:**1186-91.

71. Weixlbaumer A, Murphy FV, Dziergowska A, Malkiewicz A, Vendeix FAP, Agris PF, Ramakrishnan V: **Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines.** *Nat Struct Mol Biol* 2007, **14:**498-502.

72. Kurata S, Weixlbaumer A, Ohtsuki T, Shimazaki T, Wada T, Kirino Y, Takai K, Watanabe K, Ramakrishnan V, Suzuki T: **Modified uridines with C5-methylene substituents at the first position of the tRNA anticodon stabilize U.G wobble pairing during decoding.** *J Biol Chem* 2008, **283:**18801-11.

73. Vendeix FAP, Murphy FV, Cantara WA, Leszczyńska G, Gustilo EM, Sproat B, Malkiewicz A, Agris PF: **Human tRNA(Lys3)(UUU) is pre-structured by natural modifications for cognate and wobble codon binding through keto-enol tautomerism.** *J Mol Biol* 2012, **416:**467-85.

74. Krüger MK, Pedersen S, Hagervall TG, Sørensen MA: **The modification of the wobble base of tRNAGlu modulates the translation rate of glutamic acid codons in vivo.** *J Mol Biol* 1998, **284:**621-31.

75. Näsvall SJ, Chen P, Björk GR: **The wobble hypothesis revisited: uridine-5-oxyacetic acid is critical for reading of G-ending codons.** *RNA* 2007, **13:**2151-64.

76. Johansson MJO, Esberg A, Huang B, Björk GR, Byström AS: **Eukaryotic wobble uridine modifications promote a functionally redundant decoding system.** *Mol Cell Biol* 2008, **28:**3301-12.

77. Ikeuchi Y, Kimura S, Numata T, Nakamura D, Yokogawa T, Ogata T, Wada T, Suzuki T, Suzuki T: **Agmatine-conjugated cytidine in a tRNA anticodon is essential for AUA decoding in archaea.** *Nat Chem Biol* 2010, **6:**277-82.

78. Murphy FV, Ramakrishnan V: **Structure of a purine-purine wobble base pair in the decoding center of the ribosome.** *Nat Struct Mol Biol* 2004, **11:**1251-2.

79. Leontis NB, Westhof E: **Geometric nomenclature and classification of RNA base pairs.** *RNA* 2001, **7:**499-512.

80. Fernández IS, Ng CL, Kelley AC, Wu G, Yu Y, Ramakrishnan V: **Unusual base pairing during the decoding of a stop codon by the ribosome.** *Nature* 2013, **500:**107-10.

F1000Prime RECOMMENDED

81. Karijolich J, Yu Y: **Converting nonsense codons into sense codons by targeted pseudouridylation.** *Nature* 2011, **474:**395-8.

F1000Prime RECOMMENDED

82. Egli M: **The steric hypothesis for DNA replication and fluorine hydrogen bonding revisited in light of structural data.** *Acc Chem Res* 2012, **45:**1237-46.

83. Khade PK, Shi X, Joseph S: **Steric complementarity in the decoding center is important for tRNA selection by the ribosome.** *J Mol Biol* 2013, **425:**3778-89.

F1000Prime RECOMMENDED

84. Noller HF: **How does the ribosome sense a cognate tRNA?** *J Mol Biol* 2013, **425:**3776-7.

85. Colominas C, Luque FJ, Orozco M: **Tautomerism and Protonation of Guanine and Cytosine. Implications in the Formation of Hydrogen-Bonded Complexes.** *J Am Chem Soc* 1996, **118:**6811-21.

86. Ninio J: **A semi-quantitative treatment of missense and nonsense suppression in the strA and ram ribosomal mutants of Escherichia coli. Evaluation of some molecular parameters of translation in vivo.** *J Mol Biol* 1974, **84:**297-313.

87. Lovmar M, Ehrenberg M: **Rate, accuracy and cost of ribosomes in bacterial cells.** *Biochimie* 2006, **88:**951-61.

88. Johansson M, Lovmar M, Ehrenberg M: **Rate and accuracy of bacterial protein synthesis revisited.** *Curr Opin Microbiol* 2008, **11:**141-7.

89. Josephson K, Hartman MCT, Szostak JW: **Ribosomal synthesis of unnatural peptides.** *J Am Chem Soc* 2005, **127:**11727-35.

F1000Prime RECOMMENDED

90. Freidinger RM: **Using the ribosome to synthesize peptidomimetics.** *F1000 Biol Rep* 2009, **1**:(53).

91. Wang L, Brock A, Herberich B, Schultz PG: **Expanding the genetic code of Escherichia coli.** *Science* 2001, **292:**498-500.

F1000Prime RECOMMENDED

92. Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T: **Cell-free translation reconstituted with purified components.** *Nat Biotechnol* 2001, **19:**751-5.

93. Ohuchi M, Murakami H, Suga H: **The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus.** *Curr Opin Chem Biol* 2007, **11:**537-42.

94. Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW: **Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome.** *Nature* 2010, **464:**441-4.

F1000Prime RECOMMENDED

95. Cochella L, Green R: **Fidelity in protein synthesis.** *Curr Biol* 2005, **15:**R536-40.

96. Hammes GG, Chang Y, Oas TG: **Conformational selection or induced fit: a flux description of reaction mechanism.** *Proc Natl Acad Sci USA* 2009, **106:**13737-41.