

Structure of a purine-purine wobble base pair in the decoding center of the ribosome

Frank V Murphy IV & V Ramakrishnan

Here we report the crystal structures of I•C and I•A wobble base pairs in the context of the ribosomal decoding center, clearly showing that the I•A base pair is of an I_{anti} • A_{anti} conformation, as predicted by Crick. Additionally, the structures enable the observation of changes in the anticodon to allow purine-purine base pairing, the 'widest' base pair geometry allowed in the wobble position.

Degeneracy of the genetic code, whereby one amino acid is specified by multiple codons that vary only in the identity of the third base of the triplet, led to the formulation of the 'wobble hypothesis' by Crick¹. These rules, which describe the base pairs allowed in the third position of the codon-anticodon minihelix, have been elaborated upon in the intervening decades², with most changes adding previously unidentified base modifications (Table 1). As the structure of codon-anticodon pairing in the 30S ribosomal subunit has been crystallographically determined³, it is possible to obtain direct information about the structural basis of the wobble rules.

Inosine was first identified as a component of tRNA by Holley *et al.*⁴, and is formed from the hydrolytic deamination of adenine in tRNA transcripts⁵. Inosine structurally resembles guanine, but lacks the 2-amino group. As a result, inosine forms base pairs with U and C as guanine does, and additionally base-pairs with A, which guanine cannot do because of a 2-amino group that is incompletely solvated in the context of the decoding center. Inosine, therefore, has a greater repertoire of potential base pairs than either adenine (primarily U) or guanine (C and U). The modification of adenine to inosine has been found for all the four-codon boxes except tRNA^{Gly} (ref. 2)—eight tRNAs in higher eukaryotes and seven in yeast, but only tRNA^{Arg2} in prokaryotes⁵.

In *Escherichia coli* inosine is present only in tRNA^{Arg_{1CG}}, which is the only tRNA to decode the codons CGU/C/A (underline indicates bases that base-pair with inosine). Although inosine has a greater range of potential base pairs than either adenine or guanine, I•C and I•U pairs are less stable than G•C and G•U pairs, respectively⁶. Additionally, decoding of adenosine-ending codons by inosine is inefficient: only 0.39% of codons in *E. coli* are CGA, but CGU and CGC are well represented⁷, and CGA is used nine-fold less in high-codon-bias genes (typically highly expressed genes). Despite these apparent inefficiencies, inosine is the sole decoding partner for an A-ending codon in three cases so far⁸ (including *E. coli*) and replaces adenosine in the wobble position almost universally.

There are two proposed geometries for I•A base pairs: (i) I_{anti} • A_{anti} as proposed by Crick¹, a Watson-Crick base pair 'stretched' by the purine-purine composition of an I•A pair; and (ii) I_{syn} • A_{anti} in which the inosine base is rotated relative to its ribose by 180° (χ or ϕ_{CN}) from that of a Watson-Crick base pair. The I_{syn} • A_{anti} geometry for wobble base pairs was proposed based on the presence of *syn•anti* arrangements for several purine-purine pairs in various crystal structures^{9–11}, the fact that rotation of the purine base relative to the ribose sugar is relatively unrestrained¹², and the hypothesis that the decoding center would favor a minimization of geometric strain and facilitate the formation of rare base tautomers¹³.

To identify the allowed geometries of base-pairing in the context of the decoding center, we solved the structures of an anticodon stem loop (ASL) derived from tRNA^{Arg_{1CG}} bound to the codons CGC, the 'most cognate' base for inosine, and CGA, the 'least cognate'. Thorough examination of these two structures reveals the conformation of the bases in the base pairs, how the anticodon accommodates a 'wide' purine-purine base pair and whether a disfavored base pair in the wobble position affects closure of the 30S subunit (Fig. 1).

As expected, the I•C base pair is the same as a canonical G•C base pair except for the missing N2-O2 hydrogen bond (Fig. 2). The I•A base pair is clearly I_{anti} • A_{anti} (Fig. 2b,d) as originally predicted¹, and not the proposed alternative I_{syn} • A_{anti} ². It had been thought that the geometric constraints imposed by the decoding center would require an I_{syn} • A_{anti} wobble base pair, partly owing to restrictions on the width of the base

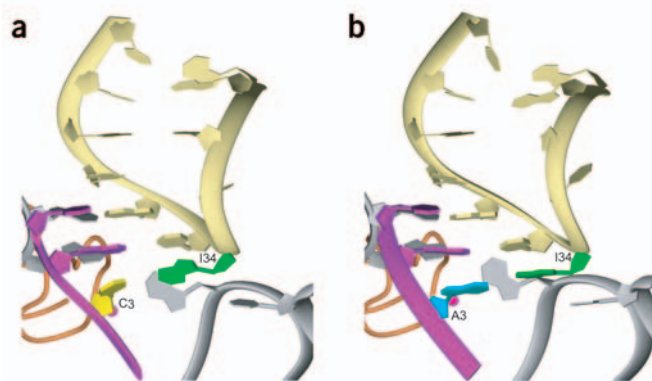


Figure 1 Cartoon structures of the I•C and I•A base pairs. (a) Overview of the decoding site for ASL^{Arg_{1CG}}•CGC. 16S rRNA is gray, with interacting bases shown, S12 is orange, codon is purple and ASL is cream. Inosine is green and cytosine is yellow. The structure has been solved to a resolution of 3.1 Å (see **Supplementary Methods**, online). (b) Overview of the decoding site for ASL^{Arg_{1CG}}•CGA. Adenine is cyan. The structure has been solved to a resolution of 3.05 Å. Atomic coordinates for ASL^{Arg_{1CG}}•CGC and ASL^{Arg_{1CG}}•CGA have been deposited in the Protein Data Bank (accession codes 1XNR and 1XNQ, respectively).

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK. Correspondence should be addressed to F.M. (fm@mrc-lmb.cam.ac.uk) or V.R. (ramak@mrc-lmb.cam.ac.uk).

Published online 21 November 2004; doi:10.1038/nsmb866

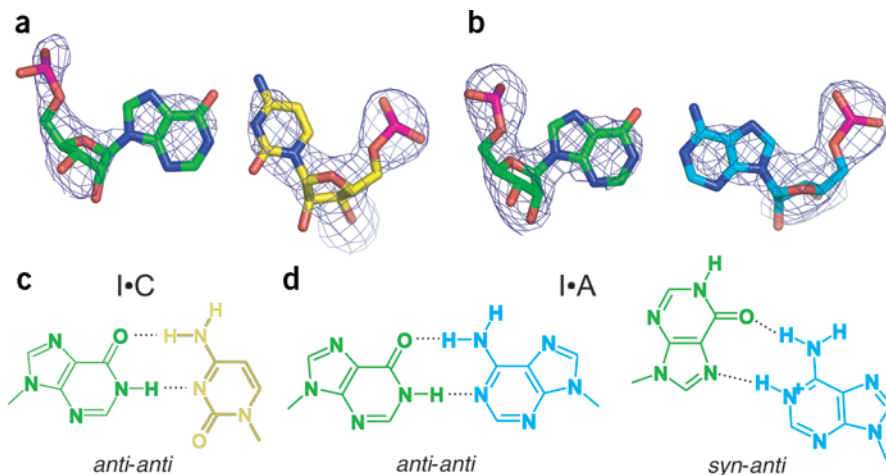


Figure 2 Details of base pairs. (a) I•C base pair (I, green; C, yellow) with $2F_o - F_c$ map contoured to 1.6σ . (b) I•A base pair (I, green; A, cyan) with $2F_o - F_c$ map contoured to 1.6σ . (c) I•C base pair with hydrogen bonds as dashed lines. (d) I•A base pairs (both *anti-anti* and *syn-anti*).

pair. However, this seems not to be the case. The decoding center places some restrictions of the position of the anticodon wobble base via C1054 (*E. coli* numbering), but the present structures show that the anticodon base is relatively free to move at the wobble position, and the base pair width is not highly constrained.

The increased width of the purine-purine I•A base pair requires the geometry of the anticodon to change, as the codon base is held in place by conserved interactions with the decoding center¹⁴. The C1'-C1' distance increases by 1.8 Å, from 10.5 to 12.3 Å (the mean C1'-C1' distance in the wobble position for all other structures so far is 10.5 ± 0.2 Å for seven structures), but the P-P distance changes only slightly, increasing by 0.2 to 1.8 Å (mean P-P distance for wobble base pairs, 17.9 ± 0.7 Å). This means that 1.6 Å of the change in base pair geometry is 'absorbed' between the C1' and P of the inosine nucleo-

side. This adjustment is made primarily by changes in the torsion angles of the main chain (β (P-O5'-C5'-C4') changes the most: -37.8° , from -177.9° to 144.3°) with very little disruption to the pucker angles of the ribose (data not shown).

The recognition of the correct base-pairing geometry in the codon-anticodon helix by the decoding center leads to a closure of the 30S subunit around the cognate tRNA; this is postulated to trigger subsequent steps in translation¹⁴. The presence of an I•A base pair in the wobble position had no substantial effect on the extent of closure of the 30S when compared with the I•C-containing structure (data not shown). Thus the decoding center recognizes an I•A base pair as cognate, with the caveat that the crystal structure observes the end point of decoding, and the kinetics of decoding could be perturbed in either the formation of the

codon-anticodon minihelix or the actual recognition of the minihelix as fully cognate.

In conclusion, the structures of I•C and I•A wobble base pairs lead to an understanding of how purine-purine base pairs at the wobble position are accommodated by the ribosome. There are other proposed *syn-anti* pairings ($G_{syn} \cdot A^+$, $G^+_{syn} \cdot G$ and $A_{syn} \cdot G$)², and in light of the evidence presented here these should be reconsidered.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

The authors thank M. Tarry for purified 30S ribosomal subunits. This work was funded by the Medical Research Council (UK), and grants from the NIH GM67624 (V.R.), and the Agouron Institute (V.R.). F.V.M. was supported by an EMBO Long-Term Fellowship.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 10 August; accepted 25 October 2004

Published online at <http://www.nature.com/nsmb/>

Table 1 Wobble rules

Anticodon	Codon
Crick's wobble rule	
G	U,C
C	G
U	A,G
I	U,C,A
Revised wobble rule	
G	U,C
C	G
k ² C	A
A	U,C,(A),G
U	U,(C),A,G
xm ⁵ s ² U, xm ⁵ Um, Um, xm ⁵ U	A,(G)
xo ⁵ U	U,A,G
I	U,C,A

k²C, lysidine; xm⁵s²U, 5-methyl-2-thiouridine derivatives; xm⁵Um, 5-methyl-2'-O-methyluridine derivatives; Um, 2'-O-methyluridine; xm⁵U, 5-methyluridine derivatives; xo⁵U, 5-hydroxyuridine derivatives. Parentheses indicate less-favored base pairs that do nonetheless form. Table is derived from ref. 2.

- Crick, F.H.C. *J. Mol. Biol.* **19**, 548–555 (1966).
- Yokoyama, S. & Nishimura, S. Modified nucleosides and codon recognition. in *tRNA: Structure, Biosynthesis and Function* (eds. Söll, D. & RajBhandary, U.) 207–223 (American Society for Microbiology, Washington, DC, 1995).
- Ogle, J.M. *et al. Science* **292**, 897–902 (2001).
- Holley, R.W., Everett, G.A., Madison, J.T. & Zamir, A. *J. Biol. Chem.* **240**, 2122–2128 (1965).
- Gerber, A.P. & Keller, W. *Science* **286**, 1146–1149 (1999).
- Grosjean, H.J., de Henau, S. & Crothers, D.M. *Proc. Natl. Acad. Sci. USA* **75**, 610–614 (1978).
- Nakamura, Y., Gojbori, T. & Ikemura, T. *Nucleic Acids Res.* **28**, 292 (2000).
- Curran, J.F. *Nucleic Acids Res.* **23**, 683–688 (1995).
- Leonard, G.A., Booth, E.D., Hunter, W.N. & Brown, T. *Nucleic Acids Res.* **20**, 4753–4759 (1992).
- Subramanian, E., Madden, J.J. & Bugg, C.E. *Biochem. Biophys. Res. Commun.* **50**, 691–696 (1973).
- Carter, R.J., Baeyens, K.J., SantaLucia, J., Turner, D.H. & Holbrook, S.R. *Nucleic Acids Res.* **25**, 4117–4122 (1997).
- Haschemeyer, A.E. & Rich, A. *J. Mol. Biol.* **27**, 369–384 (1967).
- Topal, M.D. & Fresco, J.R. *Nature* **263**, 289–293 (1976).
- Ogle, J.M., Murphy, F.V., Tarry, M.J. & Ramakrishnan, V. *Cell* **111**, 721–732 (2002).

Supplementary Methods

Thermus thermophilus 30S ribosomal subunits were purified, crystallized and cryoprotected as previously described (the final solution being 26% (v:v) MPD, 100mM K-MES pH 6.5, 200mM KCl, 75mM NH₄Cl, 15mM MgCl₂)¹. RNA oligonucleotides were chemically synthesized and gel purified (Dharmacon) with the sequences: ASL 5'-CUCGGCUICGAACCGAG-3', mRNAs 5'-CGCAAA-3', 5'-CGAAAA-3' (anticodon and codons underlined). Following cryoprotection, the 30S crystals were soaked in cryoprotection buffer containing 80 M paromomycin, 300 M ASL and 300 M of one of the mRNA hexanucleotides for at least 48 hours. Crystals were flash cooled in liquid nitrogen and stored for data collection.

Data were collected at ESRF ID14-4, processed with Denzo and Scalepack², utilizing Strategy³ to aid in the efficient collection of data. CNS 1.1⁴ was used for refinement, topologies and parameters of the modifications were taken directly or derived from HIC-Up⁵, O⁶ was used for visualization and building, the CCP4 package⁷ was used for assorted tasks, and figures were composed using PyMol⁸ and MOLMOL⁹. Alignment of the structures was performed using Escet 0.6¹⁰, which identifies structurally invariant phosphate atoms in the 16S rRNA, and analysis of RNA bond angles was performed using the program 3DNA¹¹.

References

1. Clemons, W.M., Jr. et al. Crystal Structure of the 30 S Ribosomal Subunit from *Thermus thermophilus*: Purification, Crystallization and Structure Determination. *J. Mol. Biol.* **310**, 827-843 (2001).
2. Otwinowski, Z. & Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* **276A**, 307-326 (1997).
3. Ravelli, R.B.G., Sweet, R.M., Skinner, J.M., Duisenberg, A.J.M. & Kroon, J. STRATEGY: a program to optimize the starting spindle angle and scan range for X-ray data collection. *J. Appl. Cryst.* **30**, 551-554 (1997).
4. Brunger, A.T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54** (Pt 5), 905-21 (1998).
5. Kleywegt, G.J. & Jones, T.A. Databases in Protein Crystallography. *Acta Cryst.* **D54**, 1119-1131 (1998).
6. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110-119 (1991).
7. Collaborative Computational Project, N. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763 (1994).
8. DeLano, W.L. The PyMOL Molecular Graphics System. (DeLano Scientific LLC, San Carlos, CA, 2002).
9. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51-55 (1996).
10. Schneider, T.R. A genetic algorithm for the identification of conformationally invariant regions in protein molecules. *Acta Cryst.* **D58**, 195-208 (2002).
11. Lu, X.J. & Olson, W.K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31**, 5108-5121 (2003).

Supplementary Table 1 Data collection and refinement statistics

	ASL ^{Arg} _{ICG} -CGC	ASL ^{Arg} _{ICG} -CGA
Data collection		
Space group	P4 ₁ 2 ₁ 2	P4 ₁ 2 ₁ 2
Cell dimensions		
a, b, c (Å)	401.1, 401.1, 175.9	401.1, 401.1, 175.2
α , β , γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	3.1 (3.21-3.1)	3.05 (3.16-3.05)
R_{sym} or R_{merge}	.151 (.658)	0.133(0.790)
$I/\sigma I$	14.4 (2.5)	12.8 (1.9)
Completeness (%)	96.5 (91.8)	96.6 (95.4)
Redundancy	22.2	27.7
Refinement		
Resolution (Å)	99-3.1	99-3.05
No. reflections	235669	252588
$R_{\text{work}}/R_{\text{free}}$	22.73 / 27.32	22.79 / 26.99
No. atoms	52083	52084
Protein	19237	19237
Ion	110	109
B-factors	77.01	80.06
Protein		
Ligand/ion		
Water		
R.m.s deviations		
Bond lengths (Å)	0.0068	0.066
Bond angles (°)	1.207	80.06

Highest-resolution shell is shown in parenthesis.