Specificity of RNA–RNA helix recognition

Daniel J. Battle* and Jennifer A. Doudna†‡

*Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520; and †Department of Molecular and Cell Biology and Howard Hughes Medical Institute, University of California, Berkeley, CA 94720

Edited by Ignacio Tinoco, Jr., University of California, Berkeley, CA, and approved July 9, 2002 (received for review April 12, 2002)

Functional RNAs often form compact structures characterized by closely packed helices. Crystallographic analysis of several large RNAs revealed a prevalent interaction in which unpaired adenosine residues dock into the minor groove of a receptor helix. This A-minor motif, potentially the most important element responsible for global RNA architecture, has also been suggested to contribute to the fidelity of protein synthesis by discriminating against near-cognate tRNAs on the ribosome. The specificity of A-minor interactions is fundamental to RNA tertiary structure formation, as well as to their proposed role in translational accuracy. To investigate A-minor motif specificity, we analyzed mutations in an A-minor interaction within the *Tetrahymena* group I self-splicing intron. Thermodynamic and x-ray crystallographic results show that the A-minor interaction strongly prefers canonical base pairs over base mismatches in the receptor helix, enabling RNA interhelical packing through specific recognition of Watson–Crick minor groove geometry.

Biological macromolecules must fold into compact tertiary structures to carry out catalytic activities. Whereas formation of secondary and tertiary structure is usually highly coupled in proteins, large RNAs often possess relatively stable secondary structures that can form independently of tertiary interactions. This property allows RNAs to follow a more hierarchical folding pathway in which most major secondary structural elements form first, including helices, loops, and internal bulges, followed by packing of these elements to produce the final three-dimensional structure of the molecule (1, 2). A primary aim in the study of RNA structure and folding has been to determine the rules that govern the packing of helices within an RNA secondary structure into a compact three-dimensional fold.

Recently, the A-minor motif has emerged as a universal mode of RNA helical packing, found in virtually every large RNA structure in which tertiary contacts are made (3, 4). Providing the majority of tertiary contacts used for interhelical packing in complex RNAs studied to date, the A-minor motif may be the most important element responsible for arranging RNA global architecture. A-minor interactions involve extensive contacts between unpaired adenosines extruded from one secondary structural element and the minor groove of a base pair within a receptor helix. These interactions are classified based on the position of the adenosine 2′-OH relative to the 2′-OH groups of the receptor base pair, with the first position of the receptor base pair defined as the residue closest to the 2′-OH of the adenosine (Fig. 1; refs. 3 and 4). The most common type of interaction is the type I A-minor, where both the 2′-OH and the N3 of the adenosine lie between the 2′-OH groups of the receptor base pair. Type I interactions exhibit a remarkably high degree of surface complementarity between the adenosine and the minor groove face of a Watson–Crick receptor base pair. These interactions bury an average of 195 Å^2 between the adenosine and the receptor base pair, which is the most surface area buried by any base triple (3, 5, 6). Furthermore, the adenosine is positioned by an intricate network of hydrogen bonds to both bases of the receptor pair (Fig. 1*b*).

The P4–P6 domain of the *Tetrahymena* group I intron consists of two sets of coaxially stacked helices: the independently folded P5abc subdomain and the set of coaxially stacked P4, P5, and P6 helices (P456; ref. 7). These two sets of helices are docked together by five A-minor interactions involving adenosines extruded from both an adenosine-rich bulge and a GAAA tetraloop within the P5abc subdomain (Fig. 1, ref. 3). Two involve interactions between A184 and A183 of the A-rich bulge and the C109-G212 and G110-C211 base pairs of the P4 helix, and two involve A153 and A152 of the GAAA tetraloop docked in the minor groove below P6 (the tetraloop receptor). These interactions provide the majority of the tertiary contacts between the P5abc subdomain and the P456 helical stack. The fifth A-minor base triple is the interaction between A186 and the minor groove of the P5a helix critical for the folding of the A-rich bulge and the P5abc subdomain.

Previous analysis of the type I interaction that occurs between A184 of the A-rich bulge and the C109-G212 receptor base pair revealed a strong energetic preference for adenosine (3). Similarly, the 23S rRNA phylogeny suggests a strong selection for Watson–Crick receptor base pairs, in particular C-G, over non-Watson–Crick pairs (3, 4). However, a biochemical basis for this observation has been lacking. To examine the energetic specificity associated with A-minor receptor base pairs, a mutational analysis was undertaken to investigate the C109-G212 receptor pair in the P4–P6 domain. We show here that the A-minor interaction energetically discriminates against receptor base-pair mismatches without discriminating among the various Watson–Crick pairs. These data demonstrate that the A-minor motif mediates RNA–RNA helical packing by specific recognition of Watson–Crick minor groove geometry.

Methods

Preparation of RNA Constructs. pUC19 plasmid derivatives were constructed encoding each P4–P6 construct flanked on the 5′ end by a T7 RNA polymerase promoter. Run-off transcription of these plasmids linearized with the *Ear*I restriction enzyme using T7 RNA polymerase yielded the proper RNA for crystallization experiments, which was then purified by polyacrylamide gel electrophoresis and concentrated to 6–10 mg ml^{-1} in water. RNA for folding assays was prepared by run-off transcription in the presence of [α-^{32}P]ATP or [α-^{32}P]GTP.

P4–P6 Folding Assays. P4–P6 folding assays were performed as described previously (8) except as noted below. RNA was annealed by incubation at 55°C for 7 min followed by incubation at 37°C for 10 min in a buffer containing 1 × TB (89 mM Tris base/89 mM boric acid, pH 8.5), 5% glycerol and between 0.2 and 10 mM $MgCl_2$. Native gels were prepared containing 6% acrylamide and a buffer containing 1 × TB with between 0.2 and 10 mM $MgCl_2$. Native gels were run in the same buffer at 15 Watts in a Hoefer SE 600 gel apparatus at a constant temperature of 37°C. Gels were imaged by using a Molecular Dynamics phosphorimager and the program IMAGEQUANT. The resulting data were analyzed by the program KALEIDAGRAPH. We have adopted the Hill model as the best means currently available for comparison of our results with each other as well as with

This paper was submitted directly (Track II) to the PNAS office.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.rcsb.org (PDB ID code 1L8V).

‡To whom reprint requests should be addressed. E-mail: doudna@uclink.berkeley.edu.

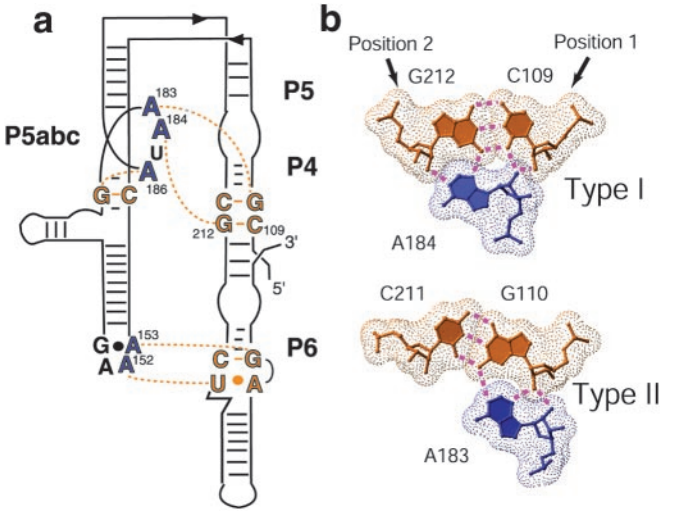

Fig. 1. The P4–P6 domain of the *Tetrahymena* group I intron is held together by five A-minor interactions. (*a*) The two primary types of A-minor interactions are identified by the position of the 2′-OH group of the adenosine relative to the 2′-OH groups of the receptor base pair. The first position of the receptor base pair is defined as the residue closest to the 2′-OH of the adenosine. The buried surface area is the area buried between the adenosine and receptor base pair. Buried surface area was calculated with the program CNS (5, 6) using a 1.4-Å radius probe and subtracting the surface area of the base triple from the sum of the surface area of the receptor base pair and the adenosine. (*b*) A-minor interactions in the P4–P6 domain. Donor adenosine residues are shown in blue, and the receptor base pairs are orange. A-minor interactions are represented by dashed lines.

previous work on the P4–P6 domain (8, 9). It should be noted that the model assumes that the RNA folding is a two-state transition exhibiting strong positive cooperativity. Values for the apparent Hill coefficient (n) and the apparent equilibrium magnesium concentration required for folding of one-half of the molecules ($[Mg^{2+}]_{1/2}$) were derived from the fit to a standard Hill equation, $F = [\{([MgCl_2]^n)/([Mg^{2+}]_{1/2}{}^n)\}/[1 + ([MgCl_2]^n)/([Mg^{2+}]_{1/2}{}^n)\}]$. $\Delta G_{apparent}$ was then calculated as $\Delta G_{apparent} = nRT\ln([Mg^{2+}]_{1/2})$. For the wild-type P4–P6 domain, we obtained values of $[Mg^{2+}]_{1/2} = 0.6$ mM and $n = 4.0$, corresponding to a $\Delta G_{apparent} = -18.1$ kcal mol^{-1}. These values are in agreement with previously reported values (8, 9). $\Delta\Delta G$ values reported are the $\Delta G_{apparent}$ of the wild-type P4–P6 subtracted from the $\Delta G_{apparent}$ of the mutant. $\Delta\Delta G$ values reflect differences in the A-minor interactions as well as possible changes in the Mg^{2+} binding properties.

X-Ray Crystallography. P4–P6-GC RNA containing mutations C109G and G212C was prepared as described above. The construct also contains a deletion of A210 shown previously to improve crystallization without affecting the structure (9). The RNA was annealed at a concentration of 4 mg ml^{-1} in a buffer containing 5 mM Hepes-NaOH, pH 7.5, 5 mM NaCl, and 40–70 mM $MgCl_2$ by heating to 65°C for 10 min and allowing it to slow cool to room temperature over ≈2 h. The RNA was then mixed at a 1:1 (vol:vol) ratio with a well solution containing 22% MPD, 50 mM Na-cacodylate, pH 6.5, and 0.05 mM spermine, and allowed to equilibrate by vapor diffusion at 25°C. After 24 h, the drops were microseeded with crystals of the P4–P6-Δ210 RNA, which were obtained as described previously (9), and crystals were generally visible after another 24 h. Crystals were cryo-stabilized by step-wise transfer to a solution containing 25% MPD, 10% isopropanol, 50 mM Na-cacodylate, pH 6.5, 100 mM $MgCl_2$, and 0.1 mM spermine. The crystals diffract to 2.8 Å resolution at the X25 beamline at the National Synchrotron Light Source at Brookhaven National Laboratory. Diffraction data were processed by using the programs DENZO and SCALEPACK (10). Phases were obtained by the molecular replacement method with the program CNS (6, 11) using the P4–P6 wild-type structure as the starting model. The structure was refined against the maximum likelihood F target to an R_{free} of 29% (Table 1) with the program CNS (6), and the model was built with the program O (ref. 12; PDB entry code 1L8V).

Results

P4–P6 Folding Assays. P4–P6 RNA folding assays were conducted to assess the energetic effect of mutation of the C109-G212 receptor base pair (3, 8). In the assay, electrophoretic mobility relative to an unfolded control molecule (J5/5a-bp, Fig. 2*a*) was monitored as a function of magnesium ion concentration. As magnesium concentration is increased, the P4–P6 molecules fold and adopt a more compact tertiary structure than the J5/5a-bp control, thereby increasing their relative mobility in a native polyacrylamide gel (Fig. 2*b*). Plots of relative electrophoretic mobilities vs. magnesium ion concentration were fit to a standard Hill equation, allowing determination of the apparent equilibrium magnesium concentration required to fold one-half of the molecules ($[Mg^{2+}]_{1/2}$), the Hill coefficient (n), and the apparent free energy of folding ($\Delta G_{apparent}$).

The Adenosine Recognizes Watson–Crick Base-Pair Geometry. P4–P6 domain variants with a mutation of the C109-G212 receptor base pair to a G-C, A-U, or U-A were constructed and subjected to analysis by the folding assay described above (Fig. 2*c*). No change in the cooperativity of folding relative to that of wild-type P4–P6 RNA was observed ($n = 3.8$, 4.2, and 4.1, respectively). Mutation

Table 1. Data statistics

	Crystallographic data				
	Resolution, Å	Completeness, %*	R_{sym}, %*†	I/σ†	Redundancy
P4-P6-GC	99–2.8	87.3/65.2	5.1/16.0	33.0/8.5	7.0

Space group $P2_12_12_1$
Unit cell a = 80.455 Å, b = 129.918 Å, c = 144.626 Å, $\alpha = \beta = \gamma = 90°$

	Refinement			RMS deviation	
	Resolution, Å	R_{cryst}, %‡	R_{free}, %‡	Bonds, Å	Angles, °
P4-P6-GC	99–2.8	29.0	29.3	0.006	1.2

*Numbers after the slash refer to the highest resolution bin.
†$R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I is the observed intensity and $\langle I\rangle$ is the statistically weighted absolute intensity of multiple measurements of symmetry-related reflections.
‡$R = \Sigma\|F(obs) - k|F(calc)\|/\Sigma|F(obs)|$, R_{cryst} from the working set and R_{free} from the test set.

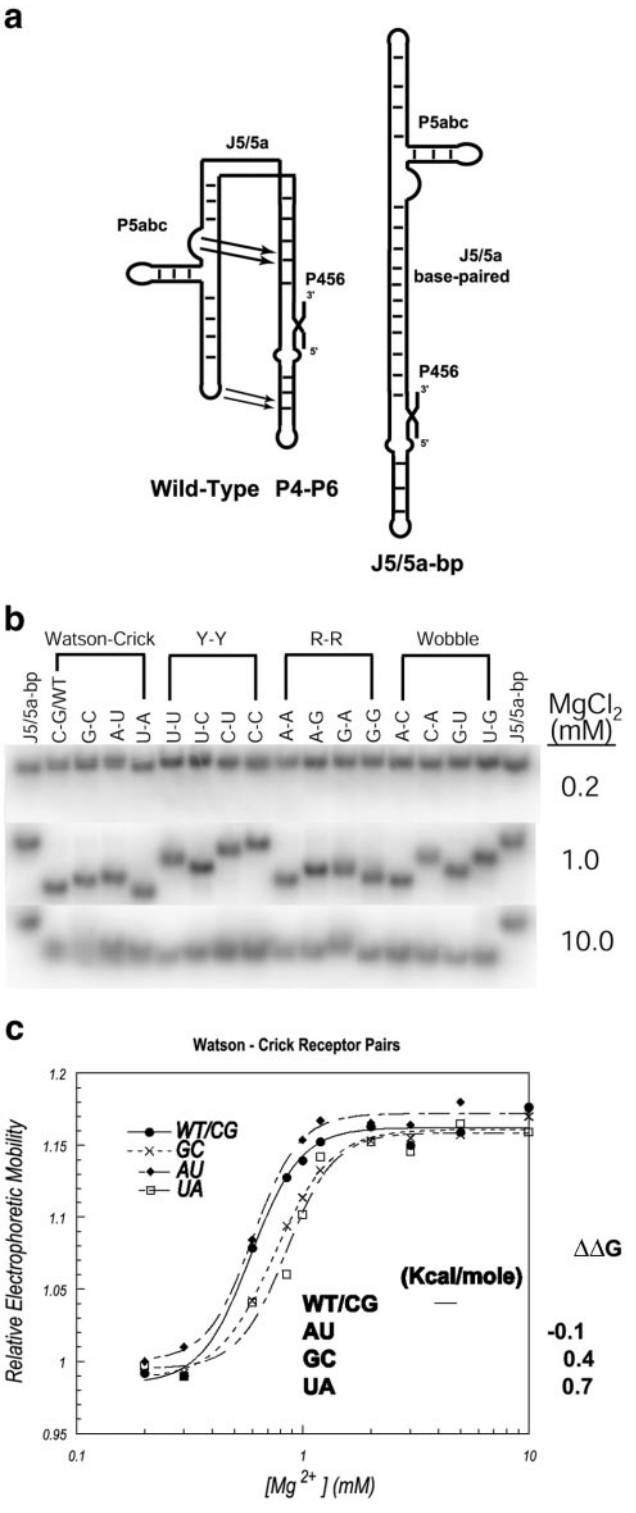

Fig. 2. P4–P6 folding assay. (*a*) Constructs used in the assay. The J5/5a-bp construct contains mutations causing the J5/5a hinge region between the P5abc subdomain and the P456 region to base pair, forming a linear molecule unable to make tertiary contacts between P5abc and P456 at any magnesium concentration (8). (*b*) P4–P6 domain variants with mutations of the C109-G212 base pair to all 16 possible base-pair combinations were incubated at various magnesium concentrations and subjected to native gel electrophoresis at constant temperature. As magnesium concentration increased, the molecules folded, resulting in increased mobility relative to the J5/5a control molecule. (*c*) Plots of electrophoretic mobility of P4–P6 domain variants with Watson–Crick base pairs at the 109–212 receptor position relative to the J5/5a-bp unfolded control molecule vs. magnesium concentration.

to an A-U base pair showed no energetic penalty relative to the wild-type ($\Delta\Delta G = -0.1$ kcal mol^{-1}). Transversion from C-G to a G-C base pair resulted in a small change ($\Delta\Delta G = 0.4$ kcal mol^{-1}), whereas mutation to a U-A base pair resulted in an only slightly larger difference ($\Delta\Delta G = 0.7$ kcal mol^{-1}). These data show that this A-minor interaction is not greatly disrupted by replacement of the wild-type C-G base pair with any other Watson–Crick base pair and therefore shows little selectivity among the various canonical pairs. This contrasts with the phylogenetic conservation observed in both group I introns and the 23S rRNA, where 99% and 61%, respectively, of the type I A-minor interactions involve C-G receptor base pairs (4, 13). Although it is unclear what accounts for the C-G receptor preference in 23S rRNA, there is evidence in group I introns that some other selective pressure on the C109-G212 position constrains its identity, such as a possible interaction of its major groove within the core of the intron (13). For this reason, this A-minor interaction may not have had any selective pressure to discriminate against other Watson–Crick pairs. Therefore, our data are not inconsistent with the group I intron phylogeny.

Crystal Structure of an A-Minor Variant. Whereas the ribosomal phylogeny might imply a significant structural perturbation upon mutation of the C-G receptor base pair to any other Watson–Crick pair, the small energetic differences measured in the P4–P6 folding assay predict only subtle effects. To analyze this, P4–P6 variants containing base-pair mutations at the A-minor receptor position were crystallized, and structures were determined by x-ray crystallography. Whereas several crystal forms did not diffract to a useful resolution, we determined a 2.8-Å crystal structure of a P4–P6 domain variant with a transversion of the C-G receptor base pair to a G-C (P4–P6-GC). Alignment on 150 phosphorous atoms resulted in a rms. difference between the P4–P6 structure and that of the P4–P6-GC mutant of 0.85 Å, showing that this mutant closely resembles the native P4–P6 structure. In the mutant structure, A184, the A-rich bulge, and the planarity of the A-minor base triple are unchanged (Fig. 3; ref. 14). A slight loss of buried surface area between the adenosine and the receptor base pair is observed, but the 190 Å^2 buried is still close to the 195 Å^2 average observed for type I A-minors and well above that of any other type of base-triple (ref. 3; Fig. 3*c*). The network of hydrogen bonds within the A-minor base triple is largely unchanged, although the distance between the O2′ of A184 and the N3 of G109 is likely too far (3.9 Å) for stable hydrogen bond formation (Fig. 3*c*). The loss of a single hydrogen bond and a slight loss of buried surface area are consistent with the $\Delta\Delta G$ of 0.4 kcal mol^{-1} observed in the folding assay. This is also consistent with previous folding studies on the P4–P6 domain in which hydrogen bonds involved in ribose zipper interactions integral to the A-minor interactions between P5abc and P456 were found to contribute -0.4 to -0.5 kcal/mol (8). This supports the interpretation of the RNA folding data for Watson–Crick base pairs as resulting from differences in A-minor interaction geometry rather than disruption of other regions of the P4–P6 RNA structure.

Selectivity Against Non-Watson–Crick Base Pairs. Phylogenetic analysis of the 23S rRNA revealed that 91.5% of the adenosines involved in type I A-minor interactions are found associated with Watson–Crick base pairs (3, 4). In addition, the recently proposed functional role of the A-minor interaction in translational accuracy presupposes the ability of the A-minor to discriminate against base mismatches (15). To determine the energetic effect of introducing base mismatches into an A-minor interaction, P4–P6 domain variants with mutations of the C109-G212 receptor base pair to all 12 possible base mismatches were subjected to a native gel electrophoresis folding assay (Fig. 4).

The A-minor interaction shows strong discrimination against

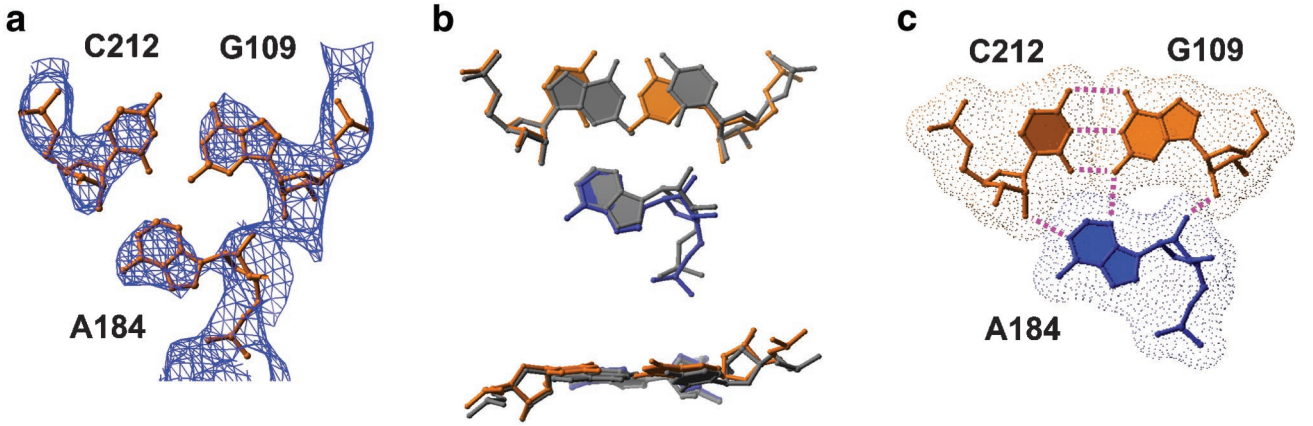

Fig. 3. Analysis of the P4–P6-GC A-minor interaction. (*a*) $2F_o - F_c$ omit map contoured at 1.2 σ in which A184, G109, and C212 were omitted from electron density calculation. (*b*) Structural alignment of the P4–P6-Δ210 (gray) and P4–P6-GC (blue and gold) structures. The P4–P6-Δ210 and P4–P6-GC structures were aligned on 150 backbone phosphorous atoms to an rms difference of 0.85 Å by using the program LSQMAN (14). (*c*) Diagram of the refined interaction highlighting the 190 Å^2 of buried surface area and the network of hydrogen bonds bridging the interaction. For comparison with the wild-type interaction, see Fig. 1.

pyrimidine–pyrimidine mismatches at the receptor base-pair position (Fig. 4*a*). The best-accommodated pyrimidine–pyrimidine base pair, U-C, still results in a 7-fold greater ΔΔG than the least stable (U-A) Watson–Crick base pair (ΔΔG = 4.8 kcal mol^{-1} vs. 0.7 kcal mol^{-1}). Larger effects are observed with the other pyrimidine–pyrimidine mismatches, with the ΔΔG of 10.8 kcal mol^{-1} for the C-C mutant being the largest observed for any base pair. In addition to the increase in $[Mg^{2+}]_{1/2}$ values, these mutations also show an effect on the cooperativity of folding, with Hill coefficients as low as 2.0 for the C-C mutant. The strong discrimination against pyrimidine–pyrimidine mismatches is likely the result of a shorter distance between the 2′-OH groups in the base pair, resulting in a narrower minor-groove face than in a Watson–Crick base pair (16). The adenosine simply cannot fit into, and is thus sterically excluded from, the minor groove. Because A183 stacks directly on A184 in the A-rich bulge, disruption of the type I interaction could possibly disrupt the type II interaction between A183 and the G110-C211 base pair immediately below it. It should also be noted that in the folding assay, some of the loss of interaction energy would be the result of destabilization of the P4 helix secondary structure by incorporation of a noncanonical base pair. Previous studies suggest that this contribution is probably small, with a predicted maximum ΔΔG of 1.3 kcal mol^{-1} (17), and therefore would not account for the majority of the discrimination observed. Additionally, we cannot rule out the possibility that the loss of the A-minor interaction disrupts the magnesium binding properties of the A-rich bulge in P5abc, although the magnesium-dependent P5abc folding occurs independently and before formation of the A-minor tertiary interactions with P456 (18, 19).

Whereas strong discrimination is observed with pyrimidine–pyrimidine mismatches, somewhat weaker discrimination is observed against purine–purine mismatches (Fig. 4*b*). The energetic penalty for introduction of a purine–purine mismatch varies from 2.1 to 3.9 kcal mol^{-1}. Although this is less than that observed for pyrimidine–pyrimidine mismatches, the ΔΔG of 2.1 kcal mol^{-1} for the A-A pair is still 3-fold greater than the largest ΔΔG observed among Watson–Crick pairs. There is only a small change in the cooperativity of folding. With these interactions, it is possible that the adenosine can still fit into the minor groove of the receptor base pair, but, because of unfavorable geometry within the minor groove pocket, it is not making the same productive interactions as seen with the Watson–Crick pairs.

Finally, mixed discrimination against wobble base pairs is observed (Fig. 4*c*). Base pairs capable of adopting wobble geometry fall into two categories depending on whether the pyrimidine of the wobble base pair is at position 2 or position 1 of the receptor pair (Fig. 4*d*). In the first case, if the pyrimidine is at position 2, then the overall effect on $\Delta G_{apparent}$ is smaller, although there is a high degree of variability. In fact, no discrimination is observed against an A-C base pair relative to a U-A Watson–Crick pair, although there is still a ΔΔG of 0.9 kcal mol^{-1} relative to the wild-type C-G pair. In addition, wobble pairs with a pyrimidine at position 2 have smaller or no changes in the cooperativity of folding. In contrast, mutation to a wobble pair with a pyrimidine at position 1 has a large energetic effect (ΔΔG of 8.3 to 10.2 kcal mol^{-1}) as well as a large effect on the cooperativity of folding. One possible explanation for this observation is that the wobble base pairs do not have symmetrical minor grooves (Fig. 4*d*; ref. 16). Instead, the pyrimidine of the wobble pair rotates back into the minor groove, creating a concave hole in the minor groove at the position of the pyrimidine nucleotide. This rotation of the pyrimidine causes a change in the angle of the N1-C1′ bond relative to that observed with Watson–Crick base pairs, thereby pulling the ribose sugar into the minor groove. This should cause a steric clash between the pyrimidine ribose and the adenosine. If there is a pyrimidine at position 1, this clash will occur on the side of the adenosine backbone, whereas if the pyrimidine is at position 2, the clash would occur on the adenosine base, away from the backbone. It may be that a steric clash on the adenosine near the backbone will cause a change in the A-rich bulge backbone position such that both the type I and the neighboring type II interactions are disrupted. The larger effects observed for the G-U and U-G over the A-C and C-A are likely the result of a second steric clash between the exocyclic amine of the guanosine and the adenosine base (Fig. 4*d*).

Discussion

Helix packing is critical to the activity and stability of many protein and RNA macromolecules. Unlike proteins, most large RNAs are characterized by stable, independently folded secondary structures, including both single-stranded regions such as loops and internal bulges, as well as regions of double-stranded helix. The abundance of conserved single-stranded adenosines in structured RNAs (20–23) and the greater accessibility of the shallow minor groove of RNA helices correlate with the observation that interhelical packing predominantly involves a single type of tertiary interaction between the single-stranded adenosines and the minor groove of neighboring double helices,

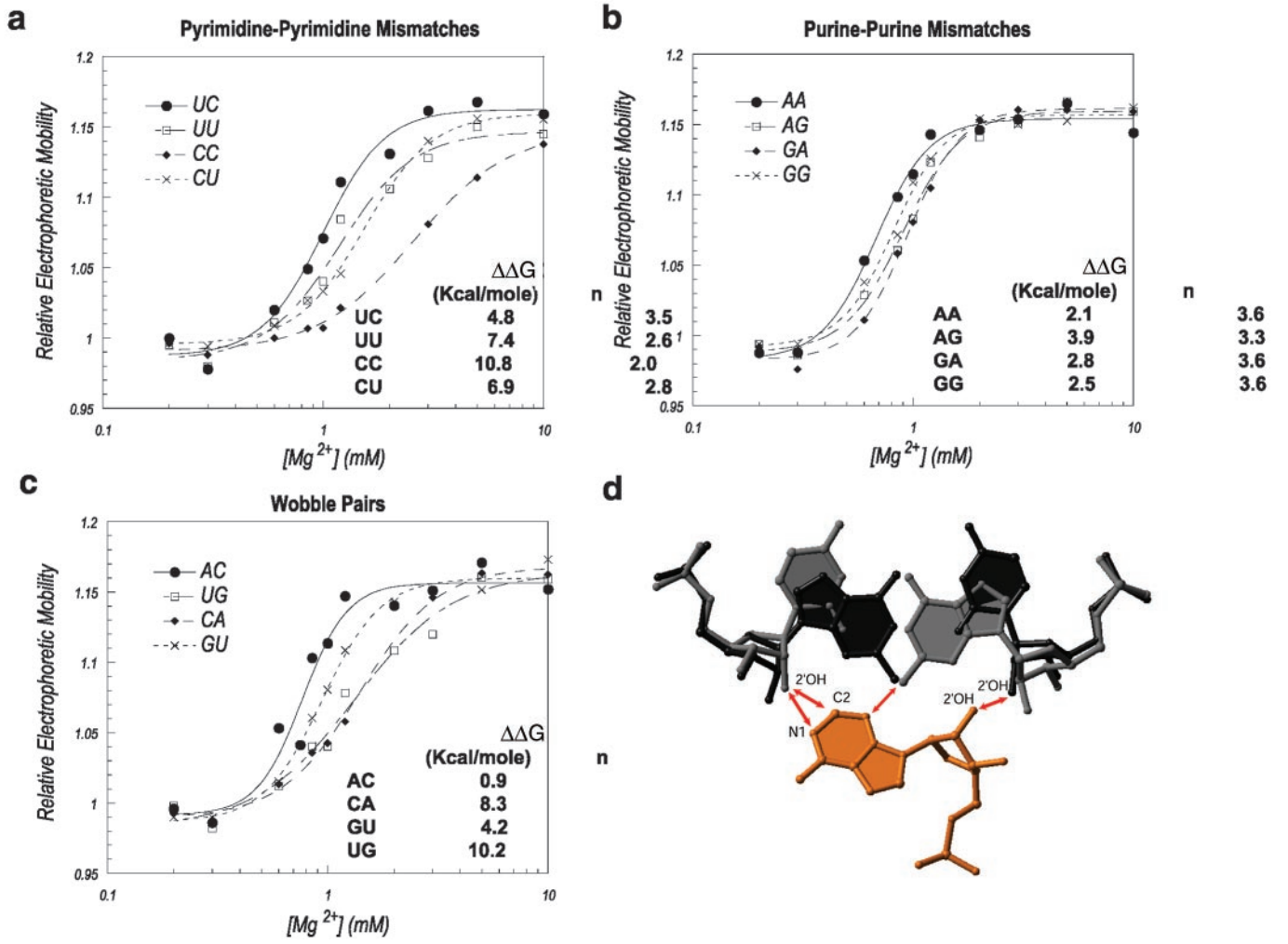

Fig. 4. Native gel folding assays conducted with P4–P6 domain variants. P4–P6 domain variants with mutations of the C109-G212 base pair to all 12 possible base mismatches were incubated at various magnesium concentrations and subjected to native gel electrophoresis at constant temperature. Shown are plots of electrophoretic mobility relative to the J5/5a-bp unfolded control molecule vs. magnesium concentration. Values for the Hill coefficient (*n*), the apparent equilibrium magnesium concentration required for folding of one-half of the molecules ($[Mg^{2+}]_{1/2}$) and $\Delta G_{apparent}$ were calculated as described. (*a*) Plots of relative electrophoretic mobility vs. magnesium concentration for P4–P6 domain variants with pyrimidine–pyrimidine mismatches at the 109–212 position. (*b*) Plots of relative electrophoretic mobility vs. magnesium concentration for P4–P6 domain variants with purine–purine mismatches at the 109–212 position. (*c*) Plots of relative electrophoretic mobility vs. magnesium concentration for P4–P6 domain variants with wobble base pairs at the 109–212 position. (*d*) Model for the effects of mutation of the receptor base pair to G-U or U-G. Wobble base pairs adapted from an oligonucleotide crystal structure (PDB entry code 485D) (16) were superimposed on the P4–P6 wild-type structure by alignment on the phosphate backbone with the program O (12). Arrows represent some likely areas of steric conflict between the model base pairs and A184.

known as the A-minor motif (3, 4). The specificity of this mode of helical docking is fundamental to the architecture of functional RNAs.

Thermodynamic measurements targeted to one of the A-minor base triples in the *Tetrahymena* group I intron P4–P6 domain show that the adenosine of a type I A-minor interaction preferentially recognizes the minor groove geometry of Watson–Crick receptor base pairs. The interaction shows little discrimination among canonical Watson–Crick pairs as measured by folding assays, and a crystal structure of a P4–P6 domain variant reveals no structural rearrangements upon transversion of the receptor base pair from C-G to G-C. On the other hand, large energetic penalties are observed for mutation of the receptor to a noncanonical base pair. The observed specificity results from the unique degree of surface complementarity between the adenosine and the minor groove face of Watson–Crick base pairs. Watson–Crick base pairs are isosteric in their minor groove geometry, and differences in the positions of individual hydrogen bond donors and acceptors in their minor grooves contribute relatively little to the selectivity. Noncanonical base pairs, however, have substantially different minor groove shapes, resulting in suboptimal fits with the adenosine. This shape-dependent selectivity for Watson–Crick minor groove geometry may form the basis for much of the interhelical packing in large RNA structures, analogous to side-chain packing between adjacent helices in proteins.

In recent years, reliable methods for the prediction of RNA secondary structure have become available (24, 25). Predicting how these secondary structural elements pack against each other to form three-dimensional globular structures has been more problematic. Because the majority of RNA helical packing is mediated by A-minor interactions, knowledge of the precise specificity and structure of the A-minor motif might, in principal, lead to ways of identifying both the adenosine nucleotides and their receptor base-pair partners involved in A-minor interactions. This would enable computational prediction of large RNA

structures, rather than the time- and labor-intensive experimental procedures currently in use. Phylogenetic analysis reveals an evolutionary preference for C-G receptor base pairs over all others (3, 4); however, the A-minor interaction studied here shows no such energetic preference. Instead, this particular A-minor interaction can accommodate any Watson–Crick base pair at the receptor position. These results show that although there may be an overall preference for a C-G pair as evidenced in the ribosomal phylogeny, system-dependent selective pressures placed on the interaction may bias the receptor pair identity for reasons not strictly related to motif stability. The findings in this study therefore demonstrate that a context-dependent component must be taken into account to predict A-minor interactions to fold large RNAs.

Many protein enzymes specifically recognize double-stranded Watson–Crick geometry by shape complementarity. DNA and RNA polymerases have been demonstrated to rely on extraordinary shape complementarity between their active site pockets and Watson–Crick base pairs to ensure fidelity in replication and transcription (26). Similarly, the specificity of the A-minor motif gives RNA the ability to select Watson–Crick base pairs by their shape. This mode of specificity may likewise allow A-minor interactions to play a functional role in macromolecular catalysis. The A-minor motif has been suggested to contribute to the fidelity of protein synthesis by discriminating against near-cognate tRNAs on the ribosome (15). In the crystal structure of the 30S ribosome complexed with tRNA and mRNA analogues, two adenosines of the 16S rRNA were observed to make A-minor interactions with the two Watson–Crick positions of the codon-anticodon helix (15). It was proposed that through these A-minor interactions, the ribosome is able to detect correct Watson–Crick geometry of the first two base pairs of the codon-anticodon helix and thereby discriminate against near-cognate tRNAs (15). To accomplish this, the A-minor interactions would have to show an energetic preference for binding to Watson–Crick base pairs without discriminating among them, which is exactly what is observed in the system studied here. Therefore, although there will be system dependencies, the observed behavior of the A-minor motif is qualitatively consistent with this proposal.

We thank Jeremy Murray, Lan Zhang, Jeff Kieft, Robert Rambo, Miguel Talavera, and the staff at National Synchrotron Light Source beamline X25 for assistance with x-ray data collection, Jeremy Murray and Andrej Luptak for comments on the manuscript, and all members of the Doudna lab for helpful discussions. This work was supported by grants from the National Institutes of Health, the National Science Foundation, and the David and Lucile Packard Foundation (to J.A.D.).

1. Sigler, P. B. (1975) *Annu. Rev. Biophys. Bioeng.* **4,** 477–527.
2. Tinoco, I., Jr., & Bustamante, C. (1999) *J. Mol. Biol.* **293,** 271–281.
3. Doherty, E. A., Batey, R. T., Masquida, B. & Doudna, J. A. (2001) *Nat. Struct. Biol.* **8,** 339–343.
4. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B. & Steitz, T. A. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 4899–4903.
5. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55,** 379–400.
6. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D* **54,** 905–921.
7. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996) *Science* **273,** 1678–1685.
8. Silverman, S. K. & Cech, T. R. (1999) *Biochemistry* **38,** 8691–8702.
9. Juneau, K., Podell, E., Harrington, D. J. & Cech, T. R. (2001) *Structure (London)* **9,** 221–231.
10. Otwinowski, Z. & Minor, W. (1996) *Methods Enzymol.* **276,** 307–326.
11. Rossman, M. G. & Blow, D. M. (1962) *Acta Crystallogr. A* **15,** 24–51.
12. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, J. (1991) *Acta Crystallogr. A* **47,** 110–119.
13. Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Muller, K. M., *et al.* (2002) *Biomed. Central Bioinform.* **3,** 2.
14. Kleywegt, G. J. (1996) *Acta Crystallogr. D* **54,** 842–857.
15. Ogle, J. M., Brodersen, D. E., Clemons, W. M., Jr., Tarry, M. J., Carter, A. P. & Ramakrishnan, V. (2001) *Science* **292,** 897–902.
16. Masquida, B. & Westhof, E. (2000) *RNA* **6,** 9–15.
17. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T. & Turner, D. H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9373–9377.
18. Murphy, F. L. & Cech, T. R. (1993) *Biochemistry* **32,** 5291–5300.
19. Sclavi, B., Sullivan, M., Chance, M. R., Brenowitz, M. & Woodson, S. A. (1998) *Science* **279,** 1940–1943.
20. Gutell, R. R., Weiser, B., Woese, C. R. & Noller, H. F. (1985) *Prog. Nucleic Acid Res. Mol. Biol.* **32,** 155–216.
21. Woese, C. R., Winker, S. & Gutell, R. R. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 8467–8471.
22. Michel, F. & Westhof, E. (1990) *J. Mol. Biol.* **216,** 585–610.
23. Gutell, R. R., Cannone, J. J., Shang, Z., Du, Y. & Serra, M. J. (2000) *J. Mol. Biol.* **304,** 335–354.
24. James, B. D., Olsen, G. J. & Pace, N. R. (1989) *Methods Enzymol.* **180,** 227–239.
25. Zuker, M. (2000) *Curr. Opin. Struct. Biol.* **10,** 303–310.
26. Kool, E. T. (1998) *Biopolymers* **48,** 3–17.