

# Wild-Type RNA Microhelix<sup>Ala</sup> and 3:70 Variants: Molecular Dynamics Analysis of Local Helical Structure and Tightly Bound Water

Maria C. Nagan, Stephanie S. Kerimo, Karin Musier-Forsyth,\* and Christopher J. Cramer\*

Contribution from the Department of Chemistry and Supercomputer Institute, University of Minnesota, 207 Pleasant St. SE, Minneapolis, Minnesota 55455-0431

Received December 11, 1998. Revised Manuscript Received June 11, 1999

**Abstract:** Molecular dynamics simulations of RNA microhelix<sup>Ala</sup> indicate that G:U and other 3:70 purine:pyrimidine wobble pairs induce local deviations from A-form geometry in their respective microhelices; the helix is underwound at the base-pair step above and overwound at the base-pair step below, in each case by about 7–9° compared to canonical A-form RNA. On the basis of analysis of average water densities and residence lifetimes, the wild-type microhelix strongly binds a water molecule in the minor groove of the 3:70 base pair, consistent with crystallographic analyses of an RNA duplex derived from the acceptor stem of *Escherichia coli* tRNA<sup>Ala</sup>. Other wobble pairs show water binding at this position but to a lesser degree; the strength of water binding correlates directly with the measured aminoacylation activities of the microhelices as substrates for *E. coli* alanyl-tRNA synthetase (G:U > 2AA:IsoC > G:dU > I:U). Watson–Crick base pairs at the 3:70 position show no tendency toward specific hydration. This tightly bound minor-groove water in the microhelices with 3:70 wobble pairs evidently does not function to stabilize a particular local helical structure, but it may play a role as a specific recognition element or serve as an indicator of interaction specificity between the microhelix and a hydrogen-bonding residue of the aminoacyl-tRNA synthetase.

## Introduction

Aminoacyl tRNA-synthetases accomplish the specific charging of individual tRNA molecules with their appropriate amino acids. Different synthetases are highly specific for their substrate tRNAs, this being a requirement for fidelity in protein synthesis. The means by which recognition is accomplished in various systems continues to be an area of active study. In the case of the alanine system, it is known that critical tRNA recognition elements reside primarily within the first three base pairs of the amino acid acceptor stem.<sup>1</sup> In particular, the structurally unique G3:U70 wobble base pair is the major recognition element in this system.<sup>2,3</sup> Recognition of acceptor stem elements by alanyl-tRNA synthetase (AlaRS) occurs even for small RNA hairpin microhelices and duplexes (which lack a closing loop) derived from the acceptor stem domain of *Escherichia coli* tRNA<sup>Ala</sup>; these smaller oligonucleotides are substrates for specific aminoacylation with alanine.<sup>4,5</sup>

Activities of duplex<sup>Ala</sup> and/or full-length tRNA<sup>Ala</sup> 3:70 variants as substrates for AlaRS have been assayed in vitro (Figure 1).<sup>6–9</sup> Most 3:70 variants, in particular G:deoxyU (G:

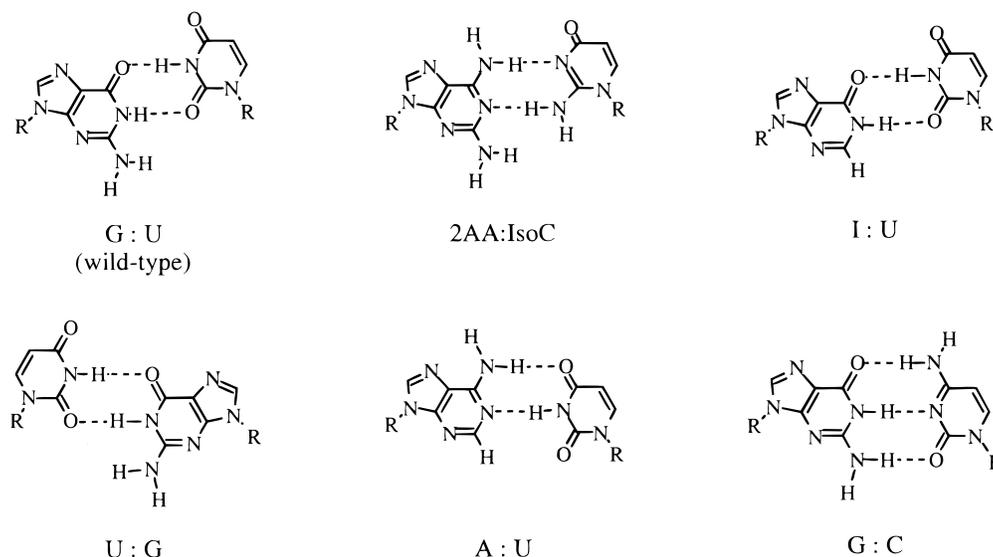
A <sup>76</sup>	
C	
C	
A	
<sup>1</sup> G-C	
G-C	
<sup>3</sup> <span style="border: 1px solid black; padding: 2px;">G-U</span> <sup>70</sup> ⇒	
G-C	
C-G	
U-A	
C-G	
U G	
U C	

variant	$\Delta\Delta G^\ddagger$ , kcal/mol
2AA:IsoC	0.6
G:dU	1.7
I:U	≥ 3.3
U:G	≥ 4.1
A:U	≥ 4.1
G:C	6.5

**Figure 1.** Wild-type RNA microhelix<sup>Ala</sup> in which the sequence of the first six base pairs and the single-stranded 3' end are identical with the acceptor stem of *E. coli* tRNA<sup>Ala</sup>. The RNA also contains a stable UUCG tetraloop closed with a C:G base pair. MD simulations were carried out for the wild-type sequence and six variants differing at the 3:70 base pair (boxed, numbering corresponds to full-length tRNA). The effect of each base pair substitution on aminoacylation catalytic efficiency is reported as  $\Delta\Delta G^\ddagger$ , which is defined as  $-RT \ln[(k_{cat}/K_M)^{variant}/(k_{cat}/K_M)^{wild-type}]$ . These values were previously determined for duplex and/or full-length tRNA<sup>Ala</sup>, as further described in refs 6–9.

dU), inosine:U (I:U), U:G, A:U, and G:C, display significantly reduced aminoacylation efficiency relative to a wild-type G:U-containing substrate: expressing these differences as differential free energies of activation, these variants show a range of increase in  $\Delta G^\ddagger$  from 1.7 to 6.5 kcal/mol compared to wild-type. The 2-aminoadenine:isocytidine (2AA:IsoC) variant, however, is aminoacylated by AlaRS with reasonable efficiency relative to wild-type ( $\Delta\Delta G^\ddagger = 0.6$  kcal/mol).<sup>8</sup> This variant exhibits wobble base pairing and moreover the 2-amino group

- (1) Musier-Forsyth, K.; Schimmel, P. *Acc. Chem. Res.* **1999**, 32, 368.
- (2) Hou, Y.-M.; Schimmel, P. *Nature* **1988**, 333, 140.
- (3) McClain, W. H.; Foss, K. *Science* **1988**, 240, 793.
- (4) Francklyn, C.; Schimmel, P. *Nature* **1989**, 337, 478.
- (5) Musier-Forsyth, K.; Scaringe, S.; Usman, N.; Schimmel, P. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, 88, 209.
- (6) Musier-Forsyth, K.; Usman, N.; Scaringe, S.; Doudna, J.; Green, R.; Schimmel, P. *Science* **1991**, 253, 784.
- (7) Musier-Forsyth, K.; Schimmel, P. *Nature* **1992**, 357, 513.
- (8) Musier-Forsyth, K.; Shi, J. P.; Henerson, B.; Bald, R.; Fürste, J. P.; Schimmel, P. *J. Am. Chem. Soc.* **1995**, 117, 7253.
- (9) Beuning, P. J.; Yang, F.; Schimmel, P.; Musier-Forsyth, K. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, 94, 10150.



**Figure 2.** Hydrogen bonding motifs for base pairs at the 3:70 position of microhelix<sup>Ala</sup>.

of 2AA occupies a position in the minor groove similar to that occupied by the amino group of G in wild-type G:U (Figure 2). An earlier study demonstrated that deletion of the minor-groove amino group of the G:U pair by substitution of I:U (Figure 2) significantly reduced aminoacylation activity.<sup>6</sup> Thus, it seems reasonable to expect the 2AA:IsoC to be a good substrate for aminoacylation. On the other hand, the guanine amino group of U:G also occupies a position in the minor groove quite similar to that occupied by the amino group in wild-type G:U (Figure 2). The large difference in activity between G:U and U:G suggests that aminoacylation is sensitive to the nature of the wobble pair and the angle at which the 2-amino group protrudes into the groove.<sup>10</sup> In particular, purine:pyrimidine (R:Y) wobble pairs, like G:U and 2AA:IsoC, are preferred over pyrimidine:purine (Y:R) wobble pairs, like U:G. Watson–Crick base pairs, like A:U and G:C, are very poor substrates for aminoacylation ( $\Delta\Delta G^\ddagger > 4.1$  kcal/mol). These base pairs either lack a minor groove amino group (A:U) or have an amino group that is hydrogen bonded (G:C).

A high-resolution NMR structure of the wild-type microhelix<sup>Ala</sup> has been described,<sup>11</sup> and X-ray crystal structures for analogous wild-type and G3:C70 variant duplexes lacking the single-stranded 3' end have also been recently reported.<sup>12</sup> In the wild-type X-ray structure, a specific network of water molecules is observed in the minor groove around the G3:U70 base pair, but this network is not present in the (inactive) G3:C70 variant. These X-ray data suggest the possibility that specific hydration may play some role in tRNA<sup>Ala</sup> recognition. Simulations have also revealed specific hydration patterns in a different tRNA G:U wobble pair, namely that found in the tRNA<sup>Asp</sup> anticodon hairpin.<sup>13</sup>

The experiments described above suggest that base substitution can affect tRNA<sup>Ala</sup> recognition by altering the functional groups exposed in the minor groove and/or by influencing the specific hydration shell structure in the vicinity of the G:U wobble pair. An additional (nonmutually exclusive) possibility is that base pair substitution may affect recognition by altering

the global or local structure of the double helix<sup>14</sup> (i.e., by introducing twisting, bending, or other features not typical of A-form RNA). In the absence of detailed structural data from experiment, simulation offers itself as an excellent tool for comparing the different variants with respect to all of the above issues. Cheatham and Brooks<sup>15</sup> and Beveridge et al.<sup>16</sup> have recently reviewed successful applications of simulation to elucidating subtle structural and environmental differences in a variety of DNA and RNA systems.

In this work, we describe molecular dynamics (MD) simulations carried out for the seven microhelix<sup>Ala</sup> variants listed in Figure 1. We focus in particular on the propensity for specific water molecules to be tightly bound in the vicinity of the 3:70 base pair for different choices of constituent bases, and on the extent to which local and global helical structure is influenced by base-pair substitution together with potentially enhanced water binding. Our goal is to find correspondences between the structural and dynamical properties of these seven microhelices, which differ only at the 3:70 base pair, and their measured aminoacylation activities.

## Computational Methods

Simulations of the wild-type G:U microhelix, and variants 2AA:IsoC, G:dU, I:U, U:G, G:C, and A:U, were all performed according to the same protocol. Initial structures for molecular dynamics simulations were constructed from a high-resolution NMR structure of the *E. coli* microhelix<sup>Ala</sup> determined by Ramos and Varani,<sup>11</sup> with nucleotide base substitutions at the 3 and/or 70 positions as warranted. The simulations employed the force field of Cornell et al.<sup>17</sup> as implemented in the AMBER 5.0 molecular dynamics suite.<sup>18</sup> Necessary parameters were developed for nucleotides not defined in the force field; parametrization procedures are described in the Supporting Information.

(14) Gabriel, K.; Schneider, J.; McClain, W. H. *Science* **1996**, *271*, 195.

(15) Cheatham, T. E., III; Brooks, B. R. *Theor. Chem. Acc.* **1998**, *99*, 279.

(16) Beveridge, D. L.; Young, M. A.; Sprous, D. In *Molecular Modeling of Nucleic Acids*; Leontis, N. B., Santa Lucia, J., Eds.; ACS Symp. Ser. No. 682; American Chemical Society: Washington, DC, 1998; p 260.

(17) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(18) Case, D. A.; Pearlman, D. A.; Cladwell, J. W.; Cheatham, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER version 5.0*; University of California: San Francisco, 1997.

(10) Frugier, M.; Schimmel, P. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11291.

(11) Ramos, A.; Varani, G. *Nucleic Acids Res.* **1997**, *25*, 2083.

(12) Mueller, U.; Schübel, H.; Sprinzl, M.; Heinemann, U. *RNA* **1999**, *5*, 670.

(13) Auffinger, P.; Westhof, E. *J. Mol. Biol.* **1997**, *269*, 326.

Microhelices were surrounded by sodium counterions and 9 Å of TIP3P<sup>19</sup> water. The particle mesh Ewald<sup>20</sup> (PME) formalism was used to account for long-range electrostatics. Prior to initializing dynamics, each microhelix was subjected to the equilibration protocol of Cheatham.<sup>21,22</sup> After this, all simulations were further equilibrated for 500 ps under production conditions prior to acquiring statistics on the trajectories. Following this, each trajectory was propagated collecting statistics for 2.0 ns; propagation was accomplished using the SANDER module of AMBER 5.0 with a 2.0 fs time step. The SHAKE algorithm<sup>23</sup> was applied to all hydrogen atoms and the nonbonded-interactions cutoff was set to 9.0 Å. The pairlist was updated every 25 steps. Constant pressure (1 bar) and temperature (300 K) were maintained according to the Berendsen<sup>24</sup> coupling algorithm.

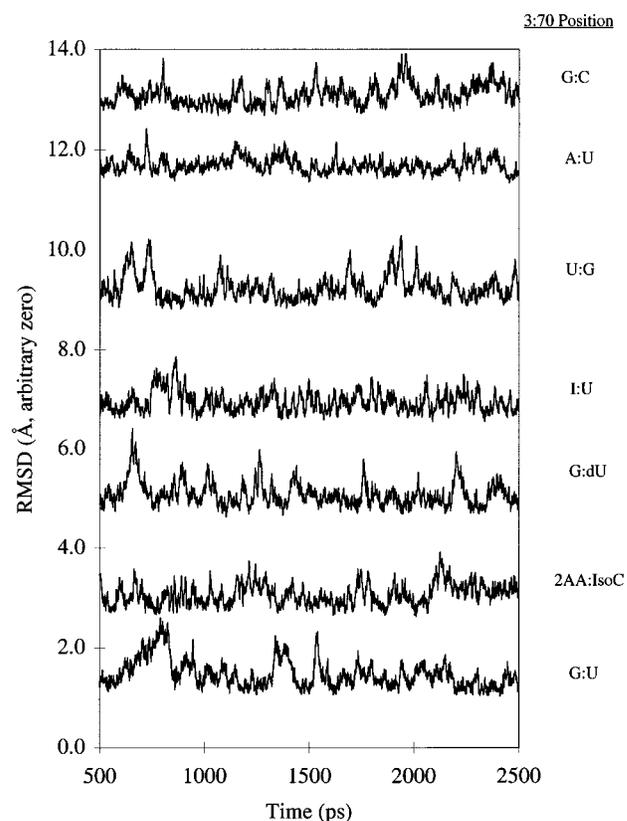
Trajectories were analyzed using the CARNAL module of AMBER 5.0. Helical parameters<sup>25,26</sup> were analyzed for snapshots taken every 1 ps using Curves 5.3.<sup>27</sup> Reported parameters, mean values, and standard deviations were computed over all 2000 snapshots.

Water contour analysis was performed for snapshots taken every 1 ps using the RDPARM module of AMBER 5.0 to generate water density grids. Probabilities are reported as the number of times in the 2000 snapshots that a water oxygen occupies a cubic grid volume 0.5 Å on a side. The cubic grid is defined relative to a single structure from the full simulation, and each snapshot is aligned with that structure prior to binning the water molecules in the snapshot. MidasPlus<sup>28</sup> and its Density delegate were used for displaying probability contours. Hydrogen bonding patterns in the vicinity of the 3:70 base pair were visualized using Insight<sup>29</sup> and VMD.<sup>30</sup>

## Results

**Convergence.** For the interpretation of data from molecular simulations to be meaningful, the modeled system must sample chemically relevant regions of phase space. By starting from the wild-type NMR structure, making nucleotide base substitutions to generate the different variants, we ensure that, to the maximum extent possible, our simulations begin *near* the most relevant region(s) of phase space. To judge the stability/convergence of each structure over its 2 ns trajectory, we analyzed the variance of several different geometrical variables.

Root-mean-square deviations from the average structures not including the ACCA single-stranded region (since it is, for the most part, disordered in solution<sup>11</sup>) were found to fluctuate in a converged fashion about characteristic mean values for each trajectory (Figure 3). Furthermore, comparisons of average structures from the first nanosecond to the second nanosecond resulted in self-symmetric root-mean-square deviations that were all smaller than 0.61 Å (Supporting Information). Many other structural parameters, e.g., local twist angle, base-pair inclination, and *x*-displacement from the global axis, were also found to oscillate with small fluctuations around essentially constant values (Supporting Information). All of these analyses are



**Figure 3.** Root-mean-square deviations from average structures taken over the entire trajectory. Root-mean-square deviation values are presented for an arbitrary zero and were calculated excluding the single-stranded region of the microhelix.

consistent with those reported for other polynucleotides judged to be well converged in the literature,<sup>15,31</sup> and we consider the seven simulations to be sufficiently stable over the analysis interval to warrant interpretation.

**Helix Structural Parameters.** Trajectories for the 7 variants were analyzed with respect to local and global helical parameters as described by Lavery et al.<sup>25,26</sup> With the exception of base twisting about the 3:70 position, all helical parameters were identical to within one standard deviation for the 7 different variants. Local twisting at the 2:71/3:70 and 3:70/4:69 base-pair steps, however, did show specific sensitivity to the nature of the 3:70 base pair. Table 1 presents the computed twist parameters determined dynamically for each of the 7 variants, as well as static equivalents determined experimentally from (i) a low-resolution NMR structure of a wild-type duplex,<sup>32,33</sup> (ii) a high-resolution NMR structure of the wild-type microhelix,<sup>11</sup> and (iii) single-crystal X-ray structures of wild-type and G:C duplexes.<sup>12</sup>

**Water Densities.** One analysis of regions of tightly bound water was accomplished by dividing space about the average structure into cubes having a volume of 0.125 Å<sup>3</sup>. Snapshots of the system taken every picosecond were aligned (microhelix only) with the average structure, and water molecules were then binned into the volume element occupied by the water oxygen atom. Over 2 ns (2000 snapshots) the average occupation of any bin, given a random box of water at standard bulk density,

(19) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(20) York, D. M.; Darden, T. A.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *99*, 8345.

(21) Cheatham, T. E., III; Kollman, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 4805.

(22) Cheatham, T. E., III *Simulation Protocol for Polynucleotides*; 1998; <http://www.amber.ucsf.edu/amber/tutorial/polyA-polyT>.

(23) Ryckaert, J.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(24) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Comput. Phys.* **1984**, *81*, 3684.

(25) Lavery, R.; Skelnar, H. *J. Biomol. Struct. Dynam.* **1988**, *1*, 63.

(26) Lavery, R.; Skelnar, H. *J. Biomol. Struct. Dynam.* **1989**, *4*, 655.

(27) Lavery, R.; Skelnar, H. *Curves version 5.3*; Paris, 1998.

(28) *MidasPlus*; Computer Graphics Laboratory, University of California: San Francisco, 1994.

(29) *Insight version 97.0*; Molecular Simulations Inc.

(30) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.

(31) Cheatham, T. E., III; Kollman, P. A. In *Structure, Motion, Interaction, and Expression of Biological Macromolecules*; Sarma, R. H., Sarma, M. H., Eds.; Adenine Press: New York, 1998; p 99.

(32) Limmer, S.; Reif, B.; Ott, G.; Lubos, A.; Sprinzl, M. *FEBS Lett.* **1996**, *385*, 15.

(33) Vogtherr, M.; Schübel, H.; Limmer, S. *FEBS Lett.* **1998**, *429*, 21.

**Table 1.** Local Twist Angles (deg) above and below the 3:70 Base Pair

base-pair step	R:Y wobble						Watson–Crick			Y:R wobble
	G:U NMR	G:U X-ray <sup>c</sup>	G:U MD <sup>d</sup>	G:dU MD	I:U MD	2AA:IsoC MD	A:U MD	G:C MD	G:C X-ray <sup>c</sup>	U:G MD
2:71/3:70	20, <sup>a</sup> 19.0 <sup>b</sup>	22.6, 26.3	17.0 ± 4.1	18.2 ± 4.2	16.0 ± 4.4	19.2 ± 4.0	25.2 ± 4.1	24.7 ± 4.0	30.7, 30.9	31.6 ± 5.1
3:70/4:69	46, <sup>a</sup> 44.7 <sup>b</sup>	40.7, 35.0	39.5 ± 3.9	39.5 ± 3.9	40.0 ± 4.1	39.1 ± 3.7	30.4 ± 4.0	31.3 ± 4.0	30.2, 33.7	23.3 ± 5.1

<sup>a</sup> Reference 33. <sup>b</sup> Our analysis of the average structure from ref 11. <sup>c</sup> Reference 12. Values are reported for each of two unique duplexes found in the unit cell. <sup>d</sup> All MD values are reported as mean value ± one standard deviation over 2 ns simulation.

should be 8.4. Figure 4 indicates the volumes enclosed by a contour interval representing occupation numbers of 40+ for wild-type and five different 3:70 base pairs: G:U, 2AA:IsoC, G:dU, I:U, U:G, and G:C. This analysis reveals a region of tightly bound water positioned so as to permit hydrogen bonding to the 2-amino group of G3 and either the 2'-OH or 2-oxo group of U70 in the wild-type helix. A minor groove water molecule is also observed in a similar location in the other helices containing R:Y wobble pairs at the 3:70 position (2AA:IsoC, G:dU, and I:U). In contrast, insignificant volumes of water bound at this contour level are seen for helices containing U:G or G:C at the 3:70 position.

Further analysis of the 3:70/4:69 region of the wild-type microhelix reveals a region of somewhat less tightly bound water (appearing at the 30+ contour level) that corresponds to a water molecule that is hydrogen bonded both to the water at the 3:70 position as well as to the 2-amino group of G4. Figure 5 provides an expanded view of the wild-type 3:70/4:69 region and shows a typical snapshot illustrating the hydrogen bonding pattern for the two specifically bound water molecules. Water molecules specifically bound to both the 4:69 and 3:70 positions are also observed in a single-crystal X-ray structure of the wild-type microhelix, as discussed further below.

**Water Residence Lifetimes.** As an alternative measure of specific hydration, water molecule residence lifetimes in the 3:70 region were analyzed over each trajectory. A water molecule was defined to be "resident" if it engaged in a specific pattern of hydrogen bonding with the microhelix in the 3:70 region. For this purpose, a single hydrogen bond was defined as a heavy-atom–heavy-atom separation no larger than 4 Å and an X–H–Y angle (X, Y = heteroatom) deviating from linear by no more than one radian. Different specific patterns were monitored for each variant. For wild-type G:U, the water molecule had to be observed hydrogen bonding both to the 2-amino group of G3 and to either the 2'-hydroxyl or the 2-oxo group of U70. For the I:U variant, the water molecule had to be observed hydrogen bonding to both the 2-oxo group of U70 and the 2'-hydroxyl of U70. For the G:dU variant, the water molecule had to be observed hydrogen bonding to both the 2-oxo group of U70 and the 2-amino group of G3. In the 2AA:IsoC variant, the water molecule had to be observed hydrogen bonding to both the 2-amino group of 2AA and the 2'-hydroxyl of IsoC. In the G:C variant, the water molecule had to be observed hydrogen bonding only to the 2'-hydroxyl. The residence lifetime for any water molecule was defined as the amount of time over which it continuously satisfied the above hydrogen bonding criteria, subject to the exception that for cases where two hydrogen bonding interactions were monitored, a water molecule was still considered to be resident if for no more than 5 ps only *one* of the two interactions was maintained. Figure 6 shows the distribution of residence lifetimes for individual water molecules summing these in order of size as they contribute to the total (cumulative) simulation time. Thus, the intersection of each curve with the abscissa gives the total amount of time that *no* water molecule was considered resident,

the next point adds to that the total amount of time water molecules with the *shortest* measured lifetime were present, the subsequent point accumulates the time from the *next* shortest lifetime, etc. The wild-type microhelix, then, shows the shortest amount of time with no resident water molecules, and exhibits the water molecule resident for the longest time in any simulation (537 ps). In contrast, the G:C variant has no water molecule resident for about 87% of the simulation, and the longest time a water molecule *is* resident is under 50 ps.

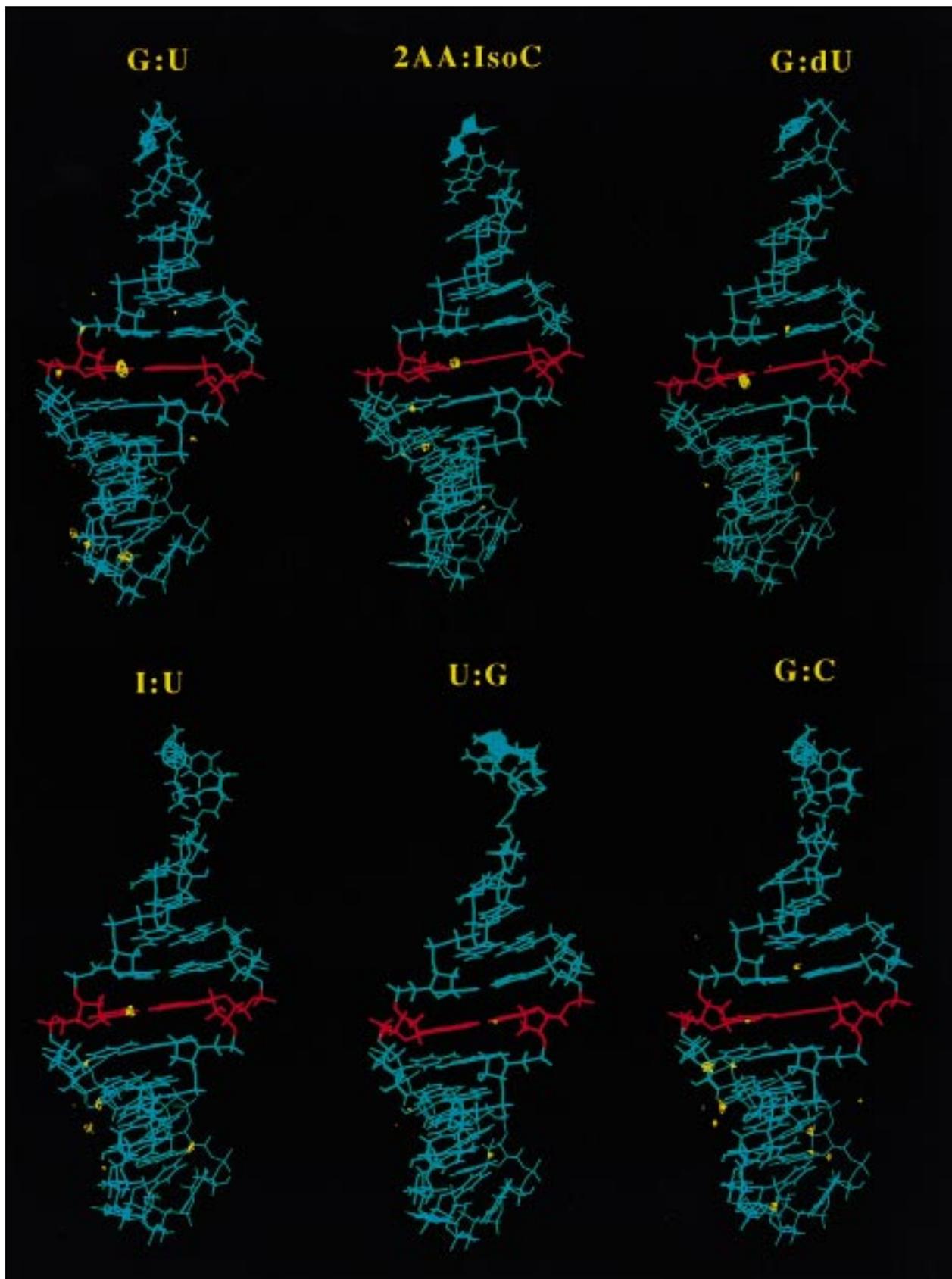
## Discussion

This section begins with a comparison of the helical geometries of the different variants. Next, the propensity of the different variants to strongly interact with specific water molecules in the vicinity of the 3:70 base pair is analyzed. Water binding, as measured by density analysis and monitoring of residence lifetimes, is discussed with respect to the degree that it may play a role in inducing/stabilizing specific helical structures, and also with respect to how it correlates with measured aminoacylation activities for the variants when they are used as substrates for AlaRS. Finally, we will speculate on the possible roles that water may play in synthetase/substrate recognition.

**Helical Twisting.** Analysis<sup>25,26</sup> of our simulations of the wild-type microhelix and its 6 variants, as well as of the average (i.e., static) high-resolution NMR structure of Ramos and Varani<sup>11</sup> for the wild-type microhelix and of the two single-crystal X-ray structures of acceptor stem duplexes,<sup>12</sup> indicates that there are no significant differences between the various sequences with respect to global helical parameters. Our observation is consistent with equivalent prior analyses, one<sup>33</sup> carried out for three low-resolution NMR structures<sup>32,33</sup> of (wild-type) G3:U70, I3:U70, and G3:C70 duplex<sup>Ala</sup> variants, and one carried out for only the two single-crystal X-ray structures<sup>12</sup> of acceptor stem duplexes, G3:U70 vs G3:C70.

Local helical parameters are similarly insensitive to sequence, with the key exception of local interbase twisting<sup>25,26</sup> above and below the 3:70 base pair. Focusing first on the wild-type G:U microhelix, the simulated result for the 2:71/3:70 twist,  $17.0 \pm 4.1^\circ$ , is in good accord with experimental values of  $19.0^\circ$  for the high-resolution NMR structure of exactly the same RNA hairpin,<sup>11</sup> and  $20^\circ$  for a low-resolution NMR structure of a wild-type duplex.<sup>32,33</sup> The single-crystal X-ray structures for G:U duplex (there are two unique duplexes in the unit cell) show somewhat larger twists of 22.6 and  $26.3^\circ$ .<sup>12</sup> All of these values represent underwinding by comparison to canonical A-form RNA.<sup>34</sup> This situation reverses for the 3:70/4:69 base-pair step. Simulation predicts an *overwinding* relative to canonical A form with a twist angle of  $39.5 \pm 3.9^\circ$ . Experiment provides similar values of 44.7, 46, 40.7, and  $35.0^\circ$  from the high<sup>11</sup> and low<sup>32,33</sup> resolution NMR structures and the single-crystal X-ray structures,<sup>12</sup> respectively.

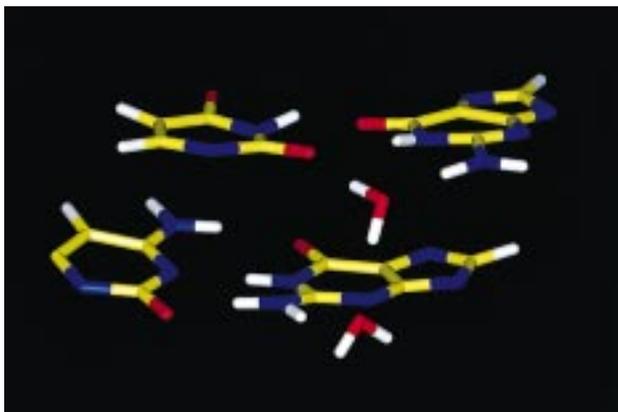
(34) Arnott, S.; Hukins, D. W. L.; Dover, S. D. *Biochem. Biophys. Res. Commun.* **1972**, *48*, 1392.



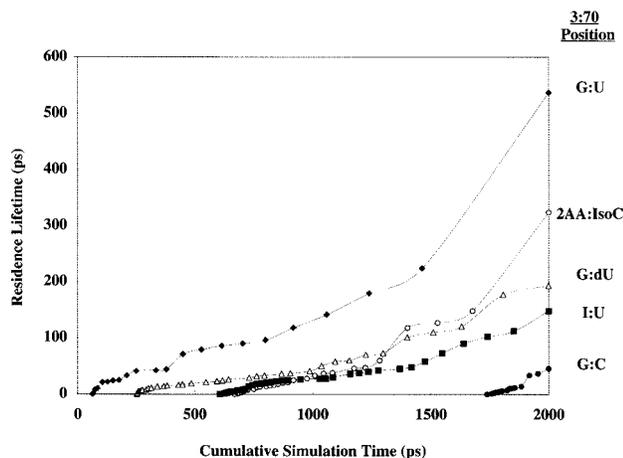
**Figure 4.** Water probability contours (40+ level, base pair 3:70 in red) for simulated (top, left to right) wild-type microhelix<sup>Ala</sup> and 3:70 variants 2AA:IsoC and G:dU, and (bottom, left to right) I:U, U:G, and G:C. Note that this view, which highlights the minor groove, places position 70 on the left and position 3 on the right, which is the reverse of the standard convention employed in Figure 2.

The underwinding/overwinding pattern observed for the G:U microhelix is maintained in the three other variants also having an R:Y wobble pair at the 3:70 position, namely G:dU, I:U,

and 2AA:IsoC. The twisting magnitudes are quite similar across the four R:Y variants, ranging from 16.0 to 19.2° for the 2:71/3:70 base-pair step and from 39.1 to 40.0° for the 3:70/4:69



**Figure 5.** Typical snapshot of the two water molecules hydrating the 3:70 and 4:69 positions of wild-type microhelix<sup>Ala</sup>.



**Figure 6.** Water molecule residence lifetimes for 5 variants. Each point not zero on the ordinate represents the observed residence lifetime of an individual water molecule. The points are ordered left to right by increasing lifetime and plotted as they contribute cumulatively to the total simulation time.

base-pair step. The somewhat greater variation observed for the 2:71/3:70 base-pair step may derive from its proximity to the single-stranded ACCA region, which is for the most part structureless in the simulations and unlikely to sample phase space in a converged fashion over even very long-time trajectories. This may also explain the larger 2:71/3:70 twist angles observed in the X-ray crystal structures compared to the simulations and the NMR structures: the RNA duplexes analyzed by X-ray crystallography lack the ACCA-3' end while it is present in all other cases.

The X-ray crystal structures<sup>12</sup> for the G:C variant duplex (again there are two duplexes in the unit cell) exhibit local twisting parameters about the 3:70 position that are entirely consistent with canonical A-form structure, with values near 31° for each base-pair step. The simulations for A:U and G:C also give twist angles of near 31° for the 3:70/4:69 base-pair step, but are underwound by about 6° for the 2:71/3:70 base-pair step. Again, we suggest that this difference probably can be attributed to the presence of the ACCA single-stranded region in the simulations but not in the crystal structures.

Finally, when the 3:70 position is occupied by a Y:R wobble pair, as is the case for U:G, the same deviations from canonical A-form structure as with R:Y base pairs are observed, but in the opposite sense (as is expected<sup>33,35</sup>). If we take the Watson-Crick simulations as the standard (thereby ensuring all com-

parisons include the single-stranded region) it is apparent that introduction of an R:Y wobble pair underwinds the 2:71/3:70 base-pair step by about 7° and overwinds the 3:70/4:69 base-pair step by about 8°, while introduction of a Y:R wobble pair overwinds the 2:71/3:70 base-pair step by about 7° and underwinds the 3:70/4:69 base-pair step by about 8°.

Only microhelices with 3:70 R:Y wobble pairs are efficiently aminoacylated by AlaRS (Figure 1). Although the presence of an R:Y wobble pair appears to be necessary, it is not sufficient for efficient charging. For example, an I:U-containing duplex is a poor substrate, despite the fact that the helix has a very similar local geometry as the wild-type RNA (Table 1).<sup>32</sup> Additionally, in the context of full-length tRNA<sup>Ala</sup>, the presence of G3:A70 and C3:A70 mismatched pairs resulted in large (approximately 100- and 3000-fold, respectively) reductions in catalytic efficiency,<sup>9</sup> despite displaying partial activity in an in vivo assay.<sup>14</sup> Thus, in addition to a local helical distortion, specific atomic groups in and around the 3:70 base pair are important factors in determining aminoacylation kinetic efficiency.<sup>6-9</sup> A recent melting temperature analysis of microhelix<sup>Ala</sup> 3:70 variants showed that the order of stability at pH 7.0 was G:C > A:U > U:G > G:U.<sup>36</sup> The inverse relationship between stability and aminoacylation activity led Meroueh and Chow to suggest that RNA helix stability might also contribute to AlaRS recognition.<sup>36</sup> Another interesting possibility, suggested by Mueller et al. based on their X-ray crystal structural analysis,<sup>12</sup> is that the local distortion caused by a wobble pair creates a pocket that favors strong binding of one or more water molecules, and that these water molecules themselves are critical to recognition/binding. We carried out a water analysis of our microhelix variants to further address this issue, and to extend the previous X-ray analysis by examining the role of specific base and backbone functional groups on the propensity for the RNA helix to bind water tightly.

**Water Densities.** In the wild-type microhelix (G:U), Figure 4 indicates that there is a well-defined region of water density at about 5 times background level in a minor-groove position between the G3 and U70 nucleotides. Visualization of the surrounding water indicates that this region typically encapsulates a single water molecule, and that when it is occupied by a water molecule that water serves as a hydrogen-bond acceptor from the 2-amino group of guanine and as a hydrogen-bond donor to either the 2'-hydroxyl group (about 67% of the time) or the 2-oxo group (about 33% of the time) of U70. Specific hydration of G:U wobble pairs has been observed before in RNA simulations,<sup>13</sup> and experimentally,<sup>12,37-40</sup> and Westhof has suggested that such situations may be designed to enhance the hydrogen bond donating ability of bound water molecules.<sup>41</sup>

As illustrated in Figure 5, about two-thirds of the time another water molecule binds just below the first in a pattern quite similar to that observed by X-ray crystallography<sup>12</sup> of a G:U duplex<sup>Ala</sup>. This agreement between simulation and crystal structure further supports the validity of analyzing water densities from the simulations.

A probability of finding a water molecule at the 3:70 position is observed in the 2AA:IsoC variant that is similar to that for

(36) Meroueh, M.; Chow, C. S. *Nucleic Acids Res.* **1999**, *27*, 1118.

(37) Holbrook, S. R.; Cheong, C.; Tinoco, I.; Kim, S. H. *Nature* **1991**, *353*, 579.

(38) Betzl, C.; Lorenz, S.; Furste, J. P.; Bald, R.; Zhang, M.; Schneider, T. R.; Wilson, K. S.; Erdmann, V. A. *FEBS Lett.* **1994**, *351*, 159.

(39) Cruse, W. B. T.; Saludjian, P.; Biala, E.; Strazewski, P.; Prangé, T.; Kennard, O. *Biochemistry* **1994**, *33*, 4160.

(40) Biswas, R.; Wahl, M. C.; Ban, C.; Sundaralingam, J. *J. Mol. Biol.* **1997**, *267*, 1149.

(41) Westhof, E. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 125.

(35) Crick, F. H. C. *J. Mol. Biol.* **1966**, *19*, 548.

G:U (Figure 4). This result is certainly reasonable from an intuitive standpoint, since 2AA:IsoC is an R:Y wobble base pair and presents minor-groove hydrogen-bond donors and acceptors very similar to G:U. Compared to the wild-type G:U, removal of the 2'-hydroxyl in the G:dU variant or the purine 2-amino group in the I:U variant does not have much effect on the enhanced water density observed at the 3:70 position; the density contours for G:U, 2AA:IsoC, G:dU, and I:U all indicate a high probability of finding a water molecule at the 3:70 position. In the U:G variant, on the other hand, while a region of higher water density is observed at the 3:70 position, it is significantly smaller in size and magnitude than those found for the R:Y wobble pairs. Water in this region, when visualized, is observed to bind to the 2-amino group of G70 and the 2'-hydroxyl group of U3, as expected. However, relative to the rest of the helix, this position is substantially displaced from the high water density regions of the R:Y wobble pairs (Figure 4).

Water density analyses for the Watson-Crick variants A:U (data not shown) and G:C (Figure 4) reveal no regions of water probability density in the vicinity of the 3:70 base pair as significant as those observed for the R:Y wobble pairs. While water molecules can be observed in individual snapshots to hydrogen bond to the Y70 2'-hydroxyl group, this interaction does not differ from similar interactions all along the minor groove, so it is evident that specific hydration is qualitatively dependent on the nature of the local helical structure as it affects the minor groove.

Water density analysis is quite useful in a qualitative sense, and has been used previously to infer specific interactions in various polynucleotide simulations.<sup>13,21,41-43</sup> In a more quantitative sense, Shields et al. have suggested that density ratios can be treated as equilibrium constants to determine absolute free energies of water binding.<sup>42</sup> Such a prescription, however, is complicated by the dynamical nature of the analysis: motions of the RNA necessarily smear the water density when it is plotted against a single reference structure. Put more succinctly, the water densities plotted about the average structure do not necessarily resemble the water densities that would derive from, say, radial distribution analyses about particular atoms (which themselves, however, would also be sensitive to local motions of the rest of the RNA). Since the integrated free energy of binding depends on the volume of space surrounded by a specific contour level, density smearing introduces uncertainty into the quantitative analysis.

An alternative measure of the strength of a specific water-RNA interaction involves not an equilibrium measure of hydration, but a kinetic one. Thus, if one would like a better handle on the "depth" of a particular water-binding well, one can focus on how long it takes to exchange one water molecule for another within that well. This analysis too has certain ambiguities associated with it, not the least of which is that water residence lifetimes for long-lived interactions will necessarily be very slowly convergent as an ensemble property. In addition, residence lifetimes will be sensitive to characteristic diffusion times for different water models, so comparison of different models is problematic. Nevertheless, analysis of water residence lifetimes can be helpful since it provides a complementary picture to that derived from density analysis.

**Water Residence Lifetimes.** To further assess the strength of water binding, residence lifetimes for those water molecules occupying regions of significant water probability density were

examined according to the "residence" criteria defined above. Prior to discussing quantitative aspects of this analysis, we note that present day force fields have not been well tested for the accuracy of water binding, particularly to biopolymers. For instance, in six separate 500 ps simulations of the tRNA<sup>Asp</sup> anticodon loop surrounded by SPC/E water, Auffinger and Westhof<sup>13</sup> have reported residence lifetimes of up to 440 ps for water molecules in the minor groove of a G:U wobble pair (where their criterion for residence involved only hydrogen bonding to the guanine 2-amino group). However, Conte et al.<sup>44</sup> have suggested, based on analysis of experimental autocorrelation times, that water molecules cannot be considered to be long-lived unless their residence lifetimes significantly exceed 500 ps. Thus, direct comparisons between theory and experiment for specific hydration lifetimes seem currently best limited to qualitative conclusions. However, internal comparisons between different simulations (employing the same force fields and water models) may be of greater utility, and that is our focus here.

Over the combined simulations of the five different variants listed in Figure 6, the longest-lived water molecule is found in wild-type G:U (537 ps), the next longest lived in 2AA:IsoC (324 ps), the next longest lived in G:U again (224 ps), etc. An alternative way to describe the tightness with which the individual variants bind water is to consider, using only the longest-lived water molecules, how many different water molecules it takes to cover 1 ns of simulation time (note that based on our residence criteria, a water molecule will be "different" even if it has previously occupied the site but has been nonresident for more than 5 ps). This corresponds to counting the number of points on each line in Figure 6 between 1 and 2 ns. For G:U, almost a full nanosecond is covered by only 4 water molecules. For 2AA:IsoC there are 10. For G:dU the count is 11, for I:U it is 15, and for G:C the residence criterion is only satisfied for 262 ps total, making it impossible to judge the number required to fill 1 ns.

The results from the lifetime analysis are consistent with expectations based on the availability of favorable hydrogen-bonding functionality in the minor groove at the 3:70 position (Figure 2). Thus, G:U forms the most favorable binding site, with the guanine 2-amino available to act as a hydrogen-bond donor and the uracil 2-oxo group together with the uracil 2'-OH group available to act as acceptors. The 2AA:IsoC pocket replaces the uracil 2-oxo group with an IsoC 2-amino group, which congests the minor groove and diminishes binding. The G:dU and I:U variants each remove one critical hydrogen-bonding group found in the wild-type base pair, and decreased binding of water would be expected on that basis. Finally, the Watson-Crick base pair at the 3:70 position fails to expose the minor groove the way it is when an R:Y wobble pair is present, so specific long-lived water interactions appear entirely absent.

Taken together, our water analyses suggest that water is most strongly bound by wild-type G:U, somewhat less strongly by 2AA:IsoC, G:dU, and I:U, and not at all by G:C. This binding propensity parallels aminoacylation activity perfectly. Although the differentiation between water residence lifetimes in 2AA:IsoC, G:dU, and I:U may not be significant, the overall correspondence provides strong support for the notion that the capability to strongly bind a water molecule at the 3:70 position is positively correlated with aminoacylation activity. We also note that water binding evidently is *not* important in stabilizing the local helical distortions induced by the presence of a 3:70

(42) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1997**, *119*, 7463.

(43) de Winter, H.; Lescrinier, E.; Van Aerschot, A.; Herdewijn, P. *J. Am. Chem. Soc.* **1998**, *120*, 5382.

(44) Conte, M. R.; Conn, G. L.; Brown, T.; Lane, A. N. *Nucleic Acids Res.* **1996**, *24*, 3693.

wobble pair. Although the G:U, 2AA:IsoC, G:dU, and I:U variants show quite different propensities for binding water (Figures 4 and 6), they exhibit essentially identical local twisting (Table 1).

**The Role of Water.** What then is the significance of the correlation between water binding and activity? One possibility, suggested by Mueller et al.<sup>12</sup> on the basis of their crystal structure data, is that the recognition event releases the bound water, and that this entropically favorable release contributes to an overall favorable free energy of binding. As water has a total entropy of about 45 eu, with only a tiny amount deriving from the vibrational partition function, release of a completely stationary water to an ideal solution can provide up to  $-14$  kcal/mol at 310 K. This obviously represents a highly idealized limiting case, but it remains clear that substantial differences in entropies of water release from the 3:70 minor groove may be expected given significantly different restrictions on the translational and vibrational motion of that water imposed by the hydrogen-bonding functionality of the 3:70 base pair (Figure 2). The importance of water and counterion release in promoting favorable binding between proteins and nucleic acids has been emphasized previously.<sup>45–48</sup>

In addition, since the entropy of release is balanced by an unfavorable enthalpy of release (or the water molecule would not be bound in the first place), the correlation between aminoacylation activity and water molecule binding may be an indication of the nature of enthalpically favorable specific interactions present in the bound enzyme–substrate complex. For example, a hydrophilic residue of the enzyme may position a hydrogen-bonding group in a fashion analogous to that of bound water, and this interaction will contribute to the specificity of recognition. The loss of entropy for the hydrogen-bonding group is not as great as the loss of entropy to bind water because chemical energy has already been expended to assemble the protein with the hydrogen-bonding group largely constrained to a specific position. Thus, while the specific protein/RNA interaction ensures specificity (since in the absence of a favorable protein/RNA interaction the enthalpic cost of water release would not be offset), the driving force is primarily the favorable entropy of water release. This situation is entirely analogous to that for protein folding (or protein/protein complexation), where hydrogen-bonding interactions and salt bridges are thought to contribute primarily to specificity, but burial of hydrophobic surface area provides the free energy driving force.<sup>49–55</sup>

An alternative role for the bound water molecule is that it may itself be a part of the specific enzyme–substrate complex. In this case, the free energy of binding water to the RNA becomes part of the overall free energy of enzyme–substrate complexation, so it is obvious why strong water binding is important.

(45) Record, M. T., Jr.; Anderson, C. F.; Mills, P.; Mossing, M.; Roe, J. H. *Adv. Biophys.* **1985**, *20*, 109.

(46) von Hippel, P. H. *Science* **1994**, *263*, 769.

(47) Spolar, R. S.; Record, M. T., Jr. *Science* **1994**, *263*, 777.

(48) Dixit, S. B.; Jayaram, B. J. *Biomol. Struct. Dynam.* **1998**, *16*, 237.

(49) Honig, B.; Hubbell, W. L. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 5412.

(50) Serrano, L.; Horovitz, A.; Avron, B.; Bycroft, M.; Fersht, A. R. *Biochemistry* **1990**, *29*, 9343.

(51) Dao-Pin, S.; Sauer, U.; Nicholson, H.; Matthews, B. W. *Biochemistry* **1991**, *30*, 7142.

(52) Sali, D.; Bycroft, M.; Fersht, A. R. *J. Mol. Biol.* **1991**, *220*, 779.

(53) Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1994**, *3*, 211.

(54) Wimley, W. C.; Gawrisch, K.; Creamner, T. P.; White, S. H. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2985.

(55) Kangas, E.; Tidor, B. *J. Chem. Phys.* **1998**, *109*, 7522.

Although it is not possible from these simulations to be more definitive about the role of water, the data do suggest certain experiments that might be interesting. For instance, if a water molecule is required for recognition, replacing G:U with 2-hydroxyethyl-6-oxopurine:U would place the pendant hydroxyl group at roughly the position in the minor groove occupied by water in the wild-type substrate, so if water is part of the specific complex, one might expect this substrate to show aminoacylation activity; if water must be released from this position, no activity would be expected. An analogous modification of the U 2'-hydroxyl group would be similarly interesting. And, of course, computational docking studies should be undertaken as more refined models for the acceptor stem interaction domain of AlaRS become available.<sup>56</sup>

## Conclusions

Molecular dynamics simulations indicate that local deviations from A-form geometries are found in tRNA<sup>Ala</sup> microhelices when they have wobble base pairs present at the 3:70 position, e.g., the wild-type microhelix, which has a G3:U70 base pair. When a purine:pyrimidine wobble base pair is present, the helix is underwound at the base-pair step above and overwound at the base-pair step below, in each case by about  $7-9^\circ$  compared to canonical A-form RNA. No deviations from canonical A-form are observed when Watson–Crick base pairs occupy the 3:70 position. A 3:70 pyrimidine:purine wobble pair, U:G, induces twisting in the opposite sense found for purine:pyrimidine pairs, i.e., overwound above, underwound below. Correlating these results with previously measured aminoacylation activities for wild-type microhelix<sup>Ala</sup> and six different variants, it is apparent that the presence of a purine:pyrimidine wobble pair is necessary but not sufficient to ensure maximum aminoacylation catalytic efficiency. In particular, the exposure of a purine 2-amino group in the minor groove facilitates recognition and/or binding.

Further analysis of the simulation trajectories to compute average water densities and specific water molecule residence lifetimes indicates that the wild-type microhelix strongly binds a water molecule at the 3:70 base pair, and another, somewhat less strongly, at the 4:69 base pair. This result is consistent with crystallographic analyses of the acceptor stem of duplex<sup>Ala</sup>. Other purine:pyrimidine wobble pairs with different minor groove functionality show equivalent locations of water binding, but the binding is less strong; Watson–Crick base pairs at the 3:70 position show no tendency toward specific hydration. The overall strength of water binding, as measured by density and lifetime analysis, has been found to correlate positively with measured aminoacylation activities for five RNA duplexes. This tightly bound minor-groove water in the microhelices with 3:70 wobble pairs evidently does not function to stabilize a particular local helical structure, but it may play a role as a specific recognition element or serve as an indicator of interaction specificity between the microhelix and a hydrogen-bonding residue of the aminoacyl-tRNA synthetase.

**Acknowledgment.** T. Cheatham, P. Schimmel, U. Mueller, and P. Kollman provided stimulating comments. Support was derived from the Minnesota Supercomputer Institute, the NIH (GM49928), the Alfred P. Sloan Foundation, and the University of Minnesota.

**Supporting Information Available:** Simulation convergence analysis and force field parameters (PDF). This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA9842565

(56) Ribas de Pouplana, L.; Buechter, D.; Sardesai, N. Y.; Schimmel, P. *EMBO J.* **1998**, *17*, 5449.