Extraordinarily slow binding of guanosine to the *Tetrahymena* group I ribozyme: Implications for RNA preorganization and function

Katrin Karbstein and Daniel Herschlag†

Department of Biochemistry, Stanford University, Stanford, CA 94305-5307

Communicated by Gerald F. Joyce, The Scripps Research Institute, La Jolla, CA, December 9, 2002 (received for review October 25, 2002)

The *Tetrahymena* ribozyme derived from the self-splicing group I intron binds a 5′-splice site analog (S) and guanosine (G), catalyzing their conversion to a 5′-exon analog (P) and GA. Herein, we show that binding of guanosine is exceptionally slow, limiting the reaction at near neutral pH. Our results implicate a conformational rearrangement on guanosine binding, likely because the binding site is not prearranged in the absence of ligand. The fast accommodation of guanosine (10^2 to 10^3·s^{-1}) and prior structural data suggest local rather than global rearrangements, raising the possibility that folding of this and perhaps other large RNAs is not fully cooperative. Guanosine binding is accelerated by addition of residues that form helices, referred to as P9.0 and P10, immediately 5′ and 3′ to the guanosine. These rate enhancements provide evidence for binding intermediates that have the adjacent helices formed before accommodation of guanosine into its binding site. Because the ability to form the P9.0 and P10 helices distinguishes the guanosine at the correct 3′-splice site from other guanosine residues, the faster binding of the correct guanosine can enhance specificity of 3′-splice site selection. Thus, paradoxically, the absence of a preformed binding site and the resulting slow guanosine binding can contribute to splicing specificity by providing an opportunity for the adjacent helices to increase the rate of binding of the guanosine specifying the 3′-splice site.

In 1894, Fischer proposed that biological recognition could be likened to the fit between a key and lock (1). More than a century later, this view has been refined to incorporate dynamic aspects of protein structure and function. Indeed, nearly all enzymes appear to undergo some motion in the course of catalysis, often in the form of a loop or domain closure that enhances binding interactions (2). More extensive conformational reordering can occur in allosteric or signaling proteins, and recently extreme examples of proteins that may have unfolded resting states have been identified (3, 4). Whereas the modern view of protein structure encompasses these dynamics, a large database of protein structures with and without bound ligand suggests that binding sites are largely preformed in most cases and that closure, if it occurs, provides additional binding interactions (e.g., refs. 5–11).

In contrast, structural studies with many RNAs reveal extensive local reordering of the structure on ligand binding. These conformational transitions include changes in hydrogen bonding as well as base stacking to accommodate the bound ligand; e.g., binding of the arginine ligand to Tar RNA induces formation of a base triple and unstacking of residues in an internal loop to allow interaction with the arginine (12–14). The frequent observation of such structural changes on ligand binding to the characterized RNAs (≤40 nt) has led to the notion that small RNAs have great difficulty in preforming ligand-binding sites (refs. 15–18 and refs. therein).

Several observations suggested that larger RNA molecules behave more like proteins, with highly preorganized active sites. The group I, group II, and HDV ribozymes as well as tRNA (≈400, 600, 90, and 76 nt, respectively) fold cooperatively, form structures with solvent protected cores, and appear to have active site clefts akin to those observed in protein enzymes (19–26). Thus, it was suggested that larger RNAs, like proteins, could preform ligand-binding sites. This idea was seemingly supported by chemical footprinting and small-angle x-ray scattering data showing that the overall conformation of the group I ribozyme is the same with or without substrates bound (24, 25, 27, 28).

Kinetic and thermodynamic studies with group I intron-derived constructs have revealed much about how RNA behaves functionally (29–36). In addition, the catalytic reaction provides a sensitive readout for investigation of how RNA behaves structurally (37, 38). Herein we investigate a slow step of the *Tetrahymena* ribozyme reaction and show that it corresponds to binding of guanosine. The results provide evidence that the binding site for this substrate is not preformed, uncover a new mechanism for specificity, and suggest that RNA may have an inherent tendency to form local alternative interactions even in the context of a globally folded molecule.

Materials and Methods

Materials. L-21 *Sca*I ribozyme was prepared as described (36, 39). L-16 *Sca*I ribozyme has a 5-nt (GGUUU) 5′-extension relative to the L-21 *Sca*I ribozyme. It was transcribed for 30 min at 30°C with 4 mM $MgCl_2$ and 2 mM NTPs to minimize self-processing and purified via RNeasy columns (Qiagen, Chatsworth, CA). Oligonucleotides were 5′-end-labeled with [^{32}P]ATP and purified (39). 3′-End labeling was performed by 5′-end labeling 3′-AMP (Ap) to yield *Ap (*, ^{32}P label), which was ligated onto the 3′-end of $CCCUCUA_5$ with T4 RNA ligase, generating SA_5*Ap. The 3′-phosphate was removed with shrimp alkaline phosphatase. Unlabeled oligonucleotides were HPLC purified (36).

General Kinetic Methods. Unless otherwise noted, all reactions were single turnover with L-21 *Sca*I or L-16 *Sca*I ribozyme (E) in excess of trace *S or *P and were carried out at 30°C in 50 mM buffer and 10 mM $MgCl_2$. The ribozyme was folded to the active state and reactions were analyzed as described (36, 37, 40). Reactions with rate constants ≥2 min^{-1} were carried out with a rapid quench apparatus (36) (KinTek, Austin, TX).

$(k_{cat}/K_m)^G$ Values. $(k_{cat}/K_m)^G$ is the second-order rate constant for the reaction: E·S + G → products. Values were determined with four subsaturating G concentrations and E saturating with respect to S [0–10 μM G, K_d^G = 110 μM, 50–200 nM E, K_d^S = 340 pM; (ref. 36 and data not shown). Rate constants were plotted against G concentration to yield $(k_{cat}/K_m)^G$ from the slope. Fits of $(k_{cat}/K_m)^G$ values vs. pH (Eq. **1**) gave the apparent pKa, pKa_{app}, and the maximum second-order rate constant, $(k_{cat}/K_m)^G_{max}$.

$$(k_{cat}/K_m)^G_{obs} = (k_{cat}/K_m)^G_{max} \frac{10^{-pK_{a,app}}}{10^{-pK_{a,app}} + 10^{-pH}}. \qquad [1]$$

The following buffers were used: Na-acetate, pH 4.5–5.5; Na-Mes, pH 5.4–6.7; Na-MOPS, pH 6.5–7.9; Na-EPPS, pH 7.9–8.6; and Na-CHES, pH 8.6–10.1. Prior control experiments revealed no

Abbreviations: S, 5′-splice site analog, CCCUCUA or $CCCUCUA_5$; P, 5′-exon analog, CCCUCU; E, L-21 *Sca*I or L-16 *Sca*I ribozyme.

†To whom correspondence should be addressed. E-mail: herschla@cmgm.stanford.edu.

www.pnas.org/cgi/doi/10.1073/pnas.252749799

buffer-specific effects under analogous conditions (41). k_{cat}/K_m values for the G analogs UCG and GGUCG were determined analogously by using 0–5 μM UCG ($K_d^{UCG} = 37\ \mu M$) and 0–0.5 μM GGUCG ($K_d^{GGUCG} = 0.9\ \mu M$).‡

Pulse–Chase Experiments. E·SA$_5$*A·G was formed at pH 5.7 by adding saturating amounts of G to E·SA$_5$*A (0.5 mM G, 550 nM E) and waiting 10 s to allow complete G binding. Excess UCG was then added with buffer of the desired pH (Fig. 2*A*). Control experiments confirmed the efficiency of the UCG chase in preventing rebinding of G. The radiolabeled product GA$_5$*A or UCGA$_5$*A was separated by denaturing gel electrophoresis.

The dissociation rate constant for G, k_{off}^G, was calculated from fraction GA$_5$*A and the rate constant of the chemical step (Eq. **2**) at pH values that give significant dissociation and reaction of G.

$$(\text{fraction } GA_5^*A)_{pH} = \frac{k_{chem}^{pH}}{k_{chem}^{pH} + k_{off}^{G}};$$
$$k_{chem}^{pH} = \frac{k_{chem,max} \times 10^{-pK_{a,app}}}{10^{-pK_{a,app}} + 10^{-pH}}. \quad [2]$$

The value for k_{off}^G was independently determined from $pK_{a_{app}}$ from Eq. **1** and the rate constant for the chemical step. At the pH corresponding to the apparent pKa $k_{off}^G = k_{chem}$, thereby giving half-maximal $(k_{cat}/K_m)^G$. The dissociation rate constants for G from these methods agreed within 2-fold.

Determination of k_{on}^{GGUCG}. The association rate constant for GGUCG was determined in pulse–chase experiments. CCCUCUA was used instead of CCCUCUA$_5$ to allow stronger binding of UCG and GGUCG (42). E·S was preformed (0.5–2.5 μM) at 4°C with S in 10% excess over E, to ensure that essentially all E had S bound. Binding was initiated by addition of *GGUCG and stopped after specified times t_1 by 10-fold dilution with 4.5 μM unlabeled GGUCG at pH 8.1 and 30°C ($K_d^{GGUCG} = 0.4\ \mu M$). Control experiments confirmed the effectiveness of the chase in preventing rebinding of *GGUCG and in allowing fast reaction of the E·S·*GGUCG complex that had formed. Plotting the fraction of reacted *GGUCG against t_1 gives a rate constant for GGUCG binding. These rate constants are plotted as a function of E·S concentration to obtain the second-order rate constant k_{on}^{GGUCG} from the slope of a linear fit to the data.

Determination of $k_{on}^{UCGAACC}$. The association rate constant for UCGAAACC was determined analogously to k_{on}^{GGUCG}. The E·P complex was formed at pH 6.0 and 4°C (9–21 nM L-16 *Sca*I, $K_d^P < 1$ nM), with 10% excess of P. *UCGAAACC was then added to allow formation of E·P·*UCGAAACC. After varying times t_1, an excess of unlabeled UCGAAACC was added at pH 8.1, and *UCG formation was determined.

Equilibrium Binding Constants for G and G analogs. The affinities of G, GTP, UCG, and GGUCG for the E·S complex were determined from the rate constant for cleavage of S with 0–1,000 μM G, 0–200 μM GTP, 0–200 μM UCG, or 0.1–10 μM GGUCG by using a deoxyribose substitution at the cleavage site to slow reaction (31, 36).

Kinetic Simulations. Kinetic simulations were performed with BERKELEY MADONNA, Ver. 8.0.1. Values for $k_{on}^{GTP} = 0.8\ \mu M^{-1}\cdot min^{-1}$ (from the plateau of k_{cat}/K_m) and $k_{off}^{GTP} = 14\ min^{-1}$ (from $k_{off} = K_d/k_{on}$) were used. At pH 6.0, the rate constants for the forward and reverse chemical step were set to 1.0 and 0.25 min^{-1}, respectively (36), and a log-linear pH dependence with slope one (refs. 36, 41, 43, and data herein) was used to calculate rates at higher pH.

‡The guanosine residue that binds in the G site is underlined throughout for clarity.

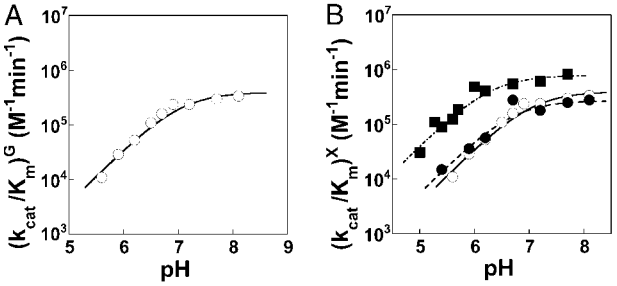


Fig. 1. The chemical step is not rate limiting at high pH values with subsaturating G and G analogs. (*A*) pH dependence of the rate constant $(k_{cat}/K_m)^G$. (*B*) The rate of the limiting step for k_{cat}/K_m is the same with G, UCG, and GGUCG. pH dependences of the rate constants $(k_{cat}/K_m)^G$ (○), $(k_{cat}/K_m)^{UCG}$ (●), and $(k_{cat}/K_m)^{GGUCG}$ (■), which describe the reactions E·CCCUCUA$_5$ + X → products. The data for $(k_{cat}/K_m)^G$ are replotted from *A* for comparison. Data were fit with Eq. **1** and give maximal values for k_{cat}/K_m of $4\cdot10^5$ $M^{-1}\cdot min^{-1}$, $3\cdot10^5$ $M^{-1}\cdot min^{-1}$, and $8\cdot10^5$ $M^{-1}\cdot min^{-1}$ and apparent pKa values of 6.7, 6.5, and 6.1 for G, UCG, and GGUCG, respectively.

Determination of $K_{1/2}^{GTP}$. Reactions were followed via rapid quench (36) by using CCCUCUA$_5$ and Mg^{2+} added stoichiometrically to GTP. A small lag in *P formation is expected from rate-limiting G binding but cannot be resolved in our experiments. The experimental and simulated time courses were therefore fit with single exponentials to test the consistency of the data with the kinetic model of rate-limiting G binding (open and closed symbols in Fig. 3*B*). Eq. **3**, derived for steady-state conditions, was also used to provide an approximate fit to both the experimental and simulated presteady state data for $K_{1/2}^{GTP}$ [pKa ≥ 10 (36)].

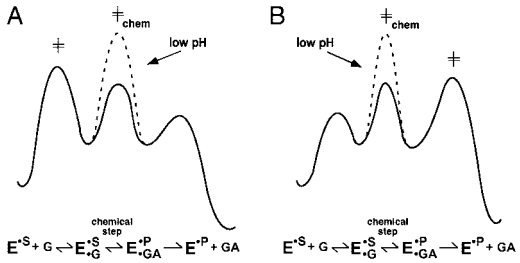
$$K_{1/2}^{pH} = K_d + \frac{k_{chem}^{pH}}{k_{on}}; \qquad k_{chem}^{pH} = \frac{k_{chem,max} \times 10^{-pK_a}}{10^{-pK_a} + 10^{-pH}}. \quad [3]$$

Results

We first present results that provide strong evidence that guanosine (G) binding is rate-limiting for $(k_{cat}/K_m)^G$ above pH 7 and then describe experiments suggesting that binding requires rearrangement of the binding site. The binding step is explored herein by using G and the tighter binding G analogs UCG and GGUCG.‡ The UC moiety of these oligonucleotides strengthens binding compared with G via formation of a helix with the ribozyme, referred to as P9.0 (42, 44, 45), and the 5′-GG residues provide additional binding energy, likely due to stacking onto the P9.0 helix (42).

Guanosine Binding Limits the Maximal Rate for $(k_{cat}/K_m)^G$. At high pH $(k_{cat}/K_m)^G$ is not limited by the chemical step. A nonchemical rate-limiting step for $(k_{cat}/K_m)^G$ had been uncovered at 50°C (43). To explore the nature of this step, we wanted to use the stronger binding of G analogs at 30°C. It was therefore necessary to confirm the presence of this rate-limiting step at 30°C. Fig. 1*A* shows that the pH dependence at 30°C levels off with an apparent pKa of 6.7 and a maximum value for $(k_{cat}/K_m)^G$ of $4\cdot10^5$ $M^{-1}\cdot min^{-1}$, similar to the maximal value of $\approx10^6$ $M^{-1}\cdot min^{-1}$ observed at 50°C (43). The reaction requires loss of a proton so the chemical step has a log-linear pH dependence with a slope of one until the pH exceeds the pKa of the 3′-OH (36, 41, 43). Because this pKa is >10 at 30°C (36), the apparent pKa of 6.7 observed here for $(k_{cat}/K_m)^G$ must be due to a change in the rate-limiting step, as is the case at 50°C (43).

The G analogs UCG and GGUCG follow a pH dependence for k_{cat}/K_m analogous to G, with similar maximal values ($5\cdot10^5$ and $8\cdot10^5$ $M^{-1}\cdot min^{-1}$, respectively; Fig. 1*B*). These results strongly suggest that reactions of G analogs are limited by the same step as reaction of G so that they can be used to help unravel the nature of the rate-limiting step above pH 7.



Scheme 1.

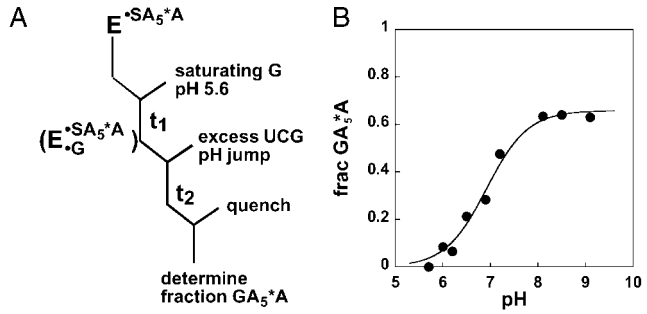


Fig. 2. At high pH, the chemical step is faster than dissociation of G. (*A*) The pulse–chase experiment. (*B*) pH dependence of the partitioning of E·S*·G between reaction and G dissociation. Data were fit with Eq. **2** and give $pKa_{app} = 7.1$.

The Rate-Limiting Step at High pH Precedes the Chemical Step. To determine the nature of the rate-limiting step above pH ≈7, we used pulse–chase experiments. These experiments delineate whether the rate-limiting step occurs before or after[§] the chemical step. If the rate-limiting step were before the chemical step (Scheme 1*A*), the reaction would occur faster than dissociation of G. Conversely, if the rate-limiting step were after the chemical step, irreversible formation of products would be slower than dissociation of G from E·S·G (Scheme 1*B*).

Because the chemical step is rate-limiting at low pH, dissociation of G is expected to be faster than reaction, leading to negligible formation of the product of reaction with G, GA_5*A (Fig. 2*A*). The fraction of GA_5*A is indeed ≈0 at low pH (Fig. 2*B*). Increasing the pH and thereby accelerating the chemical step increases the fraction of GA_5*A, indicating that more reaction with G has occurred. Thus, the rate-limiting step is before the chemical step (Scheme 1*A*).[¶] Analogous pulse–chase experiments show that the rate-limiting barrier for reaction of UCG is also before the chemical step (not shown).

If the rate-limiting step for reaction of G and UCG is before the chemical step, the principle of microscopic reversibility mandates that it be after the chemical step in the reverse reaction. To test this prediction, we followed the reverse reaction from E·P·UCGA to substrates. UCGA was used instead of GA to allow quantitative formation of the E·P·UCGA complex (36). As expected, the reverse reaction occurs with a burst in which the internal equilibrium between E·P·UCGA, and E·S·UCG is established. The reaction is then completed by a slow step after the chemical step (Fig. 6, which is published as supporting information on the PNAS web site, www.pnas.org).

Guanosine Binding Is Rate-Limiting at High pH. A rate-limiting step before the chemical step could be associated with G binding or a conformational change of the E·S·G complex. To test whether G binding was rate-limiting for $(k_{cat}/K_m)^G$, we wanted to measure the rate constant for G binding and compare it to the maximal value for $(k_{cat}/K_m)^G$. At 30°C, the affinities of G, UCG and GGUCG are too weak to readily work with ribozyme concentrations in excess of the K_d given the volumes required for rapid quench experiments. Lower temperature increases the affinity of G analogs, and at 4°C binding of the strongest-binding G analog, GGUCG, was sufficient to measure its association rate constant.

We first tested whether the kinetic pathway was maintained at the lower temperature by repeating the pH analysis for G and G analogs at 4°C. The pH dependence for k_{cat}/K_m for G, UCG, and GGUCG leveled off with apparent pKa values of 7.7–7.9 and maximal rates of 0.15, 0.35, and $2.4 \cdot 10^6$ $M^{-1} \cdot min^{-1}$, respectively (Fig. 7, which is published as supporting information on the PNAS web site), suggesting that GGUCG could be used to learn about this rate-limiting step. Pulse–chase binding experiments were carried out to determine the rate constant for GGUCG binding (Fig. 3*A*). The value of $1.4 \cdot 10^6$ $M^{-1} \cdot min^{-1}$ obtained is the same, within error, as the maximal value for $(k_{cat}/K_m)^{GGUCG}$.

To independently test the model of rate-limiting G binding, we determined the pH dependence for $K_{1/2}^{GTP}$, the GTP concentration at which the half-maximal reaction rate is observed. GTP was used instead of G because of its higher solubility. At pH ≪ 7, rapid equilibrium binding of GTP results in $K_{1/2}^{GTP}$ equal to K_d^{GTP} and independent of pH (refs. 31, 41, and 46; Fig. 3*B*). In contrast, if GTP binding is rate-limiting above pH 7 [with the chemical step log-linearly dependent on pH (36, 41, 43)], kinetic simulations predict a log-linear increase of $K_{1/2}^{GTP}$ with pH. This is analogous to the distinction between Michaelis–Menten and Briggs–Haldane steady-state kinetics, which give $K_m = K_d$ and $K_m > K_d$, respectively (47), although simulations were required here to describe the non-first-order presteady-state data. The experimental data agree quantitatively with the predicted increase in $K_{1/2}^{GTP}$ (Fig. 3*B*), providing further evidence for rate-limiting G binding above pH 7.

Given that G binding is slow relative to diffusion (see below), an unfavorable preequilibrium of the ribozyme or rate-limiting rearrangements of an intermediate with G loosely bound is implied. Kinetic data can generally not distinguish between these possibil-

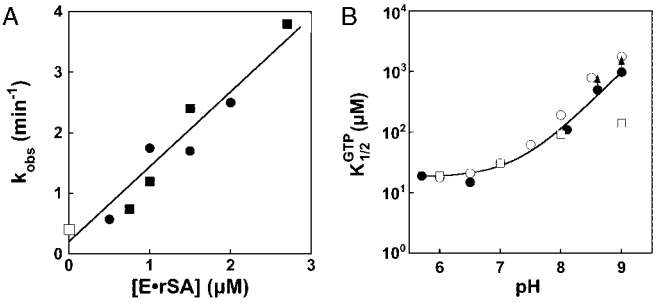


Fig. 3. Guanosine binding is rate-limiting. (*A*) Determination of the association rate constant for GGUCG, k_{on}^{GGUCG}. Data from independent experiments are shown with different filled symbols. A linear fit to the data gave $k_{on} = 1.4 \cdot 10^6$ $M^{-1} \cdot min^{-1}$ and $k_{off} = 0.2$ min^{-1}. The estimate for k_{off} of 0.3 min^{-1} (□) obtained from pKa_{app} in Eq. **1** and the rate of the chemical step is included for comparison (see *Pulse–Chase Experiments* in *Materials and Methods*). (*B*) $K_{1/2}^{GTP}$ is pH dependent. Experimental data (●) were fit to Eq. **3** and yielded $K_d^{GTP} = 19$ μM and $k_{on} = 8 (\pm 1) \cdot 10^5$ $M^{-1} \cdot min^{-1}$. The arrows indicate lower limits due to inhibition at high GTP concentrations. BERKELEY MADONNA was used to simulate GTP concentration dependences assuming rate-limiting binding (○) or a binding intermediate with 10-fold weaker binding (□). Weaker binding of G by 100-fold in the intermediate complex gives a curve essentially indistinguishable from the experimental data (not shown).

[§]A rate-limiting step can occur after the chemical step even in single-turnover reactions if the chemical conversion is thermodynamically unfavorable.

[¶]Although most of the reaction occurs with G, this fraction plateaus a maximal value of 0.68 instead of 1.0 in Fig. 2*B*. This is presumably due to a combination of factors, including a fraction of ribozyme that did not have G bound prior to the chase (18%), a fraction of ribozyme in which the S was not docked at the active site [≈2%, (50)], some damaged ribozyme (≈5%, estimated from reaction endpoints), and possibly a fraction of G that is misbound.

Table 1. Association rate constants for binding of G to E·S complexes

3′-tail of S	k_{on}^{G} (10^5 M^{-1}·min^{-1})
$_pCH_3$	5.0
$_pA$	5.8
$_pAAAAA$	4.6

The association rate constants with different E·S complexes were determined from the plateau for $(k_{cat}/K_m)^G$ at high pH values. CCCUCU$_p$Me was prepared as described (33).

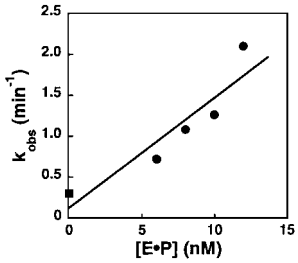


Fig. 4. The ability to form the P10 helix increases the rate of UCGAAACC binding. Data from pulse–chase experiments (see *Materials and Methods*) were fit linearly and gave $k_{on}^{UCGAAACC} = 1.4{\cdot}10^8$ M^{-1}·min^{-1} and $k_{off}^{UCGAAACC} = 0.12$ min^{-1} at 4°C. $k_{off}^{UCGAAACC} = 0.3$ min^{-1} (■) was determined independently in partitioning experiments as described in the supporting information for GAAACC.

ities unless the intermediate accumulates. Nevertheless, the pH dependence for $K_{1/2}^{GTP}$ provides a constraint on possible models: any G bound intermediate must bind G at least 10^2-fold weaker than the observed dissociation constant (Fig. 3*B*).

Origin of the Exceptionally Slow Binding of Guanosine. The results presented above provide strong evidence that G binding limits the rate of reaction of the L-21 *Sca*I ribozyme above pH 7. Remarkably, the observed rate constant for G binding of $4{\cdot}10^5$ M^{-1}·min^{-1} is 10^5-fold slower than diffusion and 10^3- to 10^4-fold slower than typically observed for association of proteins with their ligands (47). We wanted to know whether G binding was slow because the binding site was not preorganized, or whether the binding site was preformed but sterically inaccessible.

S and G bind next to each other to allow attack of the 3′-OH of G on S. The model of the group I active site places the substrate "tail," which is transferred during the reaction, in a location that could sterically restrict access of the G binding site (ref. 48; see also refs. 45 and 49).‖ To test whether the substrate tail hindered G binding, we shortened this tail to a single methyl group and measured the association rate for G. Association of G is independent of the substrate tail length (Table 1), providing strong evidence against steric hindrance from the substrate.

Most generally, steric restriction models predict that larger G analogs bind more slowly. To test this prediction, we determined the association rate constants for analogs of different size: G, UCG, CUG (which cannot form the P9.0 duplex), and GGUCG. The association rate constants for these G analogs, determined from the pH plateaus for k_{cat}/K_m, are all within 3-fold of the value for G (Fig. 1*B*; and R. Russell and D.H., unpublished results). Furthermore, UCGA, which has an additional residue 3′ to G, binds with a similar rate constant (within 3-fold; ref. 36). These data suggest that a steric block does not account for slow G binding.

It remains possible that binding of G and G analogs requires transient opening of the active site and that, once open, all G analogs bind with the same rate constant. To test this model, we determined whether the known opening transition, undocking of the P1 duplex, was required for G binding. At pH 8.6, reaction of E·S + G occurs with a rate constant of 8 s^{-1} (determined in rapid quench experiments; data not shown), whereas P1 undocking occurs at 0.2 s^{-1} [determined in single-molecule fluorescence experiments (50)]. Thus, binding and reaction of G are at least 40-fold faster than undocking of P1, indicating that transient undocking is not required. This conclusion is consistent with the previous observation that G can bind without undocking of CUCUA (51).

As described above, slow G binding could be caused by a steric block or lack of preorganization; either would render most encounters unproductive. Given the evidence against a steric block, we adopt the model in which preorganization of the G binding site is limited.

Kinetics of Rearrangements to Accommodate G. Above, we have described that G binding is slow, because most encounters with the unaligned G site are unproductive. This model implies that a conformational change is required for accommodation of G in its binding site. To learn more about this step, we used a "chemical clock" to ask how fast G was accommodated in its correct site. For this experiment, we added binding interactions in the form of base pairs 5′ and 3′ of G to lengthen encounters with the ribozyme. This increased lifetime is predicted to increase the fraction of productive binding events, giving an increase in the observed rate constant for G binding. We will first present experiments measuring binding rates of G analogs and then describe how these binding rates were used to estimate rate constants for accommodation of G into its binding site.

At 4°C, GGUCG binding is 45-fold stronger than binding of G (0.9 and 42 μM, respectively; data not shown), a result of P9.0 formation. Binding of GGUCG is also 12-fold faster than binding of G (2.4 and 0.2 μM^{-1}·min^{-1}, Fig. 7). This suggests that the P9.0 helix is formed in the transition state for binding of GGUCG,** thereby increasing the time for G accommodation and accelerating its observed binding.

The AAACC portion of (UC)GAAACC forms five base pairs with a 5′-elongated ribozyme, L-16 *Sca*I, to resemble the duplex between intron and 3′-exon, called P10, which is formed in the second step of self-splicing. The association rate constants for UCGAAACC and GAAACC are ≈100 times faster than for UCGA [2 and $3{\cdot}10^8$ M^{-1}·min^{-1} vs. $2{\cdot}10^6$ M^{-1}·min^{-1}, respectively; Fig. 4; see *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site, and ref. 36, respectively].

To test whether faster binding of (UC)GAAACC was due to an increased lifetime of encounter complexes, as described above, or instead attributable to a conformational change induced by the formation of P10, we formed P10 in *trans* and determined the association rate constant for UCG from the plateau for $(k_{cat}/K_m)^{UCG}$. Formation of P10 in *trans* did not affect the association rate constant for UCG (4 and $6{\cdot}10^5$ M^{-1}·min^{-1}, with and without P10 formed; Fig. 8, which is published as supporting information on the PNAS web site). This result suggests that formation of P10 increases the lifetime of the covalently tethered G in the vicinity of its binding site, thereby accelerating binding relative to free G.

Because P9.0 and P10 increase the rate of G binding, their

‖Whereas the exact location of the G is not well defined in current models, its position relative to S is constrained by the proximity of the 3′-OH and the phosphoryl bond (for nucleophilic attack) and through the interaction of the 2′-OH and one of the phosphoryl oxygens with the same metal ion (35). Although G could be mispositioned in the ground state, the metal ion interaction is correctly made (69). Additionally, steric hindrance observed between the adenosine 2 residues 3′ from the cleavage site (+2A) and P9.0, 5′ to G, places the substrate ''tail'' and the 5′-end of G next to each other (42).

**These results are consistent with a weakened P9.0 helix in the transition state, e.g., due to fraying base pairs or missing stacking interactions. Alternatively, P9.0 could be fully formed (giving a 45-fold effect) with G accommodation slowed 4-fold relative to the case without P9.0 formed to give the observed 12-fold effect.

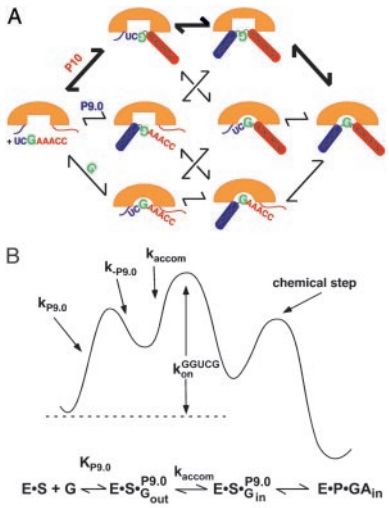


Fig. 5. Models for binding of G and G analogs. (*A*) Binding of UCGAAACC occurs in multiple steps via different pathways. Formation of P9.0 and P10 is depicted by the blue and red cylinders, respectively. Binding of G in its binding site is shown as rotation from an inverted position. Rearrangements of the binding site are represented by the change from a rectangle to a half-circle. The thicker arrows denote the kinetically favored pathway. (*B*) Schematic free energy profile for subsaturating GGUCG illustrating the binding mechanism underlying the calculation of k_{accom}. Formation and dissociation of P9.0 are shown as the barrier described by $k_{P9.0}$ and $k_{-P9.0}$, respectively. Accommodation of G in its binding site is depicted by the conversion between G_{in} and G_{out}. An analogous scheme can be considered for P10 formation (see the supporting information).

formation to give an intermediate with P9.0 and/or P10 formed must precede stable G binding, representing a distinct step before accommodation of G (Fig. 5*A*). The rate constants for (UC)GAAACC binding are similar to rate constants for duplex formation [10^8 to 10^9 $M^{-1}\cdot min^{-1}$ (52)], suggesting that formation of the P10 duplex is rate-limiting. In contrast, the rate constant for binding of GGUCG is ≈100-fold slower than duplex formation, suggesting that only 1 in 100 binding events is productive. Most simply, this suggests that G accommodation (k_{accom}, Fig. 5*B*) is ≈100-fold slower than dissociation of P9.0 ($k_{-P9.0}$; $k_{-P9.0} = 100\cdot k_{accom}$). An estimate for $k_{-P9.0}$ can be obtained by calculating the stability of P9.0 ($K_{P9.0}$) from nearest-neighbor rules [1.9–2.4 kcal/mol (53)] and the rate constant for formation of the P9.0 duplex [$k_{-P9.0} = k_{P9.0}\cdot K_{P9.0}$, with $k_{P9.0} \approx 10^8$ to 10^9 $M^{-1}\cdot min^{-1}$ (52)]. This yields an estimate of 10^4 to 10^5 s^{-1} for $k_{-P9.0}$ and, correspondingly, an estimate of 10^2 to 10^3 s^{-1} for k_{accom}. This estimate is consistent with the independently obtained lower limit of 8 s^{-1}, the rate for reaction E·S + G at high pH values described above, and the lower limit of 5 s^{-1} obtained from analogous considerations of G binding in the presence of P10 (see *Supporting Materials and Methods*).

Discussion

Limited Preorganization of the *Tetrahymena* Ribozyme Active Site. Binding of G to the group I ribozyme is 10^5-fold slower than diffusion, apparently requiring a conformational change in the ribozyme. Results with G analogs varying in size suggest that an absence of a correctly preorganized G binding site rather than steric hindrance accounts for the slow binding. Recent data on docking of the P1 duplex, in which the 5′-splice site analog is bound, similarly suggest that the binding site for P1 is not preformed (54). Further experiments probing ground-state interactions with substrates and products implicate alternative binding modes for S and UCGA (refs. 55 and 56; unpublished work). These results suggest that the active site of this RNA enzyme is not fully performed, a result that could hold generally for functional RNAs. Based on the considerations outlined below, we suggest that the ribozyme adopts an alternative local structure that must be disrupted to alllow G binding.

The differences in preorganization of protein and RNA active sites lead to the question: Why does RNA not preform ligand binding sites? As has been suggested previously (57, 58), interactions within RNA molecules are not highly specific; base pairs can form between any two bases and stacking interactions, a major driving force for nucleic acid folding, are even less specific. These properties may limit the ability of RNA to discriminate between different structures.

The global structure of the ribozyme assessed by footprinting and small angle x-ray scattering is unchanged with bound substrates (24, 25, 27, 28), whereas local rearrangements appear to be necessary for binding of substrates (data herein and ref. 54). This raises the possibility that the linkage between local and global structure in RNA molecules may be weaker than in proteins. The distinction between RNA and proteins, if general, may be due to differences in their structural organization. Most residues in a protein make tertiary interactions (59, 60), thereby forming an extensive grid of tertiary interactions. In contrast, most residues in RNA form only base pairs, with tertiary connections established by a minority of residues. This results in a structure with less interconnectivity that nevertheless retains global cooperativity in folding. These features and a less closely packed interior in a folded RNA (57) may allow more local alternatives in positioning of residues.

Slow Binding Can Increase Specificity. The results herein pose a paradox for splicing specificity. Accurate selection of the correct splice site would most simply involve stronger binding of the correct sequence at the active site. If, however, reaction of both the correct and incorrect sequences occurs faster than their dissociation, equilibration does not occur, and thermodynamic discrimination cannot be used for specificity (61). The results herein show that G binding fits this description, yet a guanosine (G414) defines the 3′-splice site. The problem then becomes how G414 is correctly selected amidst many incorrect guanosines.

The data herein show that P9.0 and P10 helices, formed 5′ and 3′ to G414, accelerate binding of G up to 10^2-fold. Thus, the correct splice site can bind faster than incorrect splice sites and will therefore react faster than incorrect ones. Evidence for the importance of kinetic selection in *Tetrahymena* self-splicing comes from observations that changing G414 alone has no influence on splice site selection, which provides evidence against thermodynamic discrimination at the G site (62). In contrast, mismatches in P10 increase selection of alternative splice sites (62).

Kinetic discrimination is possible only with a binding rate constant for G well below the physical limit imposed by diffusion. Binding at the rate of diffusion would leave no window for further acceleration. Thus, the results herein demonstrate a previously unrecognized mechanism to achieve specificity in biology: slow binding, via an induced fit, coupled with specific acceleration by cognate-binding interactions. This mechanism may be of broad importance in biology. For example, the RNA of the signal recognition particle (SRP) accelerates formation of a functional SRP/SRP receptor complex (63), and slow basal binding of major histocompatibility complexes by the T cell receptor could allow kinetic discrimination in the immune response (64).

The data herein and elsewhere (54) suggest that ligand-binding sites in the group I ribozyme are not preorganized, leading to slow binding of substrates. Further, the conformational changes required for establishing binding interactions allow increased specificity of 3′-splice site selection. Whereas it is not possible to determine from data on a single RNA whether limited preorganization is a general shortcoming or a chosen trait of RNA, we propose that RNA has learned to make the best of an inherent propensity to form alternative structures, increasing biological specificity as described above for self-splicing.

Binding Mechanism for Guanosine and Guanosine Analogs. The conformational rearrangements observed on binding of a protein or another ligand to RNA have spurred discussion as to whether association occurs through "tertiary structure capture" or "induced fit" (17, 65–67). In "tertiary structure capture," binding occurs to a sparingly populated conformer, so that an unfavorable equilibrium precedes binding. Conversely, in the "induced fit" model, rearrangements occur within an initial complex. Here, we provide evidence for a hybrid between these models, in which rearrangements occur within a complex, but through capture of local motions. These results illustrate the considerable complexity and multistep nature of binding events.

Most collisions between G and the ribozyme are unproductive. This is because the ribozyme does not have a preformed binding site, and dissociation from an encounter complex is faster than the rearrangements to establish binding interactions. As a result, G binding is slowed relative to diffusion. Binding interactions from P9.0 or P10 accelerate binding of G 10- or 100-fold, respectively. Thus, formation of P9.0 and P10 must precede stable G binding, thereby providing evidence for formation of intermediates with P9.0 or P10 formed, but G not accommodated within its binding site (Fig. 5*A*). In principle, initial binding can occur through G, P9.0, or P10, and multiple binding routes are possible. The results herein indicate that the pathway with P10 formed before P9.0 and G is kinetically favored. Indeed, formation of P10 gives a complex that dissociates slower than the conformational rearrangements to accommodate G (Fig. 9, which is published as supporting information on the PNAS web site).

Classically, such a stepwise binding mechanism falls in the induced fit category, because rearrangements occur within a complex. Nevertheless, these rearrangements, rather than actively induced by the ribozyme, could occur passively through the additional binding interactions, which allow more local motions to be sampled and individual interactions to be captured. We term this latter process "localized tertiary structure capture," and the converse scenario, in which rearrangements are facilitated, kinetically or thermodynamically, by elements within the RNA is referred to as "facilitated tertiary structure capture."

To distinguish between these binding mechanisms, we tested whether P10 formed in *trans* still provides faster binding to the G site, as could occur in facilitated tertiary structure capture, or whether tethering of P10 and G is required for the enhancement as predicted from localized tertiary structure capture. The absence of an effect from P10 in *trans* (Fig. 8) supports (UC)GAAACC binding via localized tertiary capture.

Finally, we emphasize that conformational reorganization occurs via Brownian motion, involving multiple bond rotations, so that all binding processes are complex, multistep processes (63–68). For binding of large protein and RNA ligands, the mechanisms and principles described above for binding of G and G analogs may be combined: there may be multiple intermediates that can use localized and facilitated tertiary structure capture in different steps. Furthermore, native and non-native interactions can be used to bias or facilitate subsequent steps. Unraveling these complex processes presents a fascinating future challenge.

We thank R. Russell for permission to cite unpublished data, L. Bartley for discussion, and members of the D.H. laboratory for comments on the manuscript. This work was supported by National Institutes of Health Grant GM49243 (to D.H.) and by a fellowship from the Boehringer Ingelheim Fonds (to K.K.).

1. Fischer, E. (1894) *Ber. Dt. Chem. Ges.* **27,** 2985–2993.
2. Lesk, A. M. & Chothia, C. (1984) *J. Mol. Biol.* **174,** 175–191.
3. Wright, P. E. & Dyson, H. J. (1999) *J. Mol. Biol.* **293,** 321–331.
4. Dedmon, M. M., Patel, C. N., Young, G. B. & Pielak, G. J. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 12681–12684.
5. Bode, W. & Schwager, P. (1975) *J. Mol. Biol.* **98,** 693–717.
6. Bode, W., Schwager, P. & Huber, R. (1978) *J. Mol. Biol.* **118,** 99–112.
7. James, M. N. G., Sielecki, A. R., Brayer, G. D., Delbaere, L. T. J. & Bauer, C. A. (1980) *J. Mol. Biol.* **144,** 43–88.
8. James, M. N. G., Brayer, G. D., Delbaere, L. T. J., Sielecki, A. R. & Gertler, A. (1980) *J. Mol. Biol.* **139,** 423–438.
9. Perkins, S. J., Johnson, L. N., Machin, P. A. & Phillips, D. C. (1978) *Biochem. J.* **173,** 607–616.
10. Phillips, S. E. V. (1980) *J. Mol. Biol.* **142,** 531–554.
11. Rees, D. C. & Lipscomb, W. N. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 5455–5459.
12. Puglisi, J. D., Tan, R. Y., Calnan, B. J., Frankel, A. D. & Williamson, J. R. (1992) *Science* **257,** 76–80.
13. AboulEla, F., Karn, J. & Varani, G. (1996) *Nucleic Acids Res.* **24,** 3974–3981.
14. Brodsky, A. S. & Williamson, J. R. (1997) *J. Mol. Biol.* **267,** 624–639.
15. Patel, D. J., Suri, A. K., Jiang, F., Jiang, L. C., Fan, P., Kumar, R. A. & Nonin, S. (1997) *J. Mol. Biol.* **272,** 645–664.
16. Williamson, J. R. (2000) *Nat. Struct. Biol.* **7,** 834–837.
17. Leulliot, N. & Varani, G. (2001) *Biochemistry* **40,** 7947–7956.
18. Sussman, D., Nix, J. C. & Wilson, C. (2000) *Nat. Struct. Biol.* **7,** 53–57.
19. Golden, B. L., Gooding, A. R., Podell, E. R. & Cech, T. R. (1998) *Science* **282,** 259–264.
20. Ferre, D. A. R., Zhou, K. H. & Doudna, J. A. (1998) *Nature* **395,** 567–574.
21. Costa, M. & Michel, F. (1999) *EMBO J.* **18,** 1025–1037.
22. Costa, M., Michel, F. & Westhof, E. (2000) *EMBO J.* **19,** 5007–5018.
23. Konforti, B. B., Liu, Q. L. & Pyle, A. M. (1998) *EMBO J.* **17,** 7105–7117.
24. Celander, D. W. & Cech, T. R. (1991) *Science* **251,** 401–407.
25. Wang, J. F. & Cech, T. R. (1992) *Science* **256,** 526–529.
26. Herschlag, D. (1998) *Nature* **395,** 548–549.
27. Engelhardt, M. A., Doherty, E. A., Knitt, D. S., Doudna, J. A. & Herschlag, D. (2000) *Biochemistry* **39,** 2639–2651.
28. Russell, R., Millett, I. S., Doniach, S. & Herschlag, D. (2000) *Nat. Struct. Biol.* **7,** 367–370.
29. Pyle, A. M., Murphy, F. L. & Cech, T. R. (1992) *Nature* **358,** 123–131.
30. Bevilacqua, P. C., Johnson, K. A. & Turner, D. H. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8357–8361.
31. McConnell, T. S., Cech, T. R. & Herschlag, D. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8362–8366.
32. Strobel, S. A. & Cech, T. R. (1993) *Biochemistry* **32,** 13593–13604.
33. Narlikar, G. J., Gopalakrishnan, V., McConnell, T. S., Usman, N. & Herschlag, D. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 3668–3672.
34. Narlikar, G. J. & Herschlag, D. (1998) *Biochemistry* **37,** 9902–9911.
35. Shan, S., Kravchuk, A. V., Piccirilli, J. A. & Herschlag, D. (2001) *Biochemistry* **40,** 5161–5171.
36. Karbstein, K., Carroll, K. S. & Herschlag, D. (2002) *Biochemistry* **41,** 11171–11183.
37. Russell, R. & Herschlag, D. (1999) *J. Mol. Biol.* **291,** 1155–1167.
38. Russell, R. & Herschlag, D. (2001) *J. Mol. Biol.* **308,** 839–851.
39. Zaug, A. J., Grosshans, C. A. & Cech, T. R. (1988) *Biochemistry* **27,** 8924–8931.
40. Herschlag, D. & Cech, T. R. (1990) *Biochemistry* **29,** 10159–10171.
41. Knitt, D. S. & Herschlag, D. (1996) *Biochemistry* **35,** 1560–1570.
42. Russell, R. & Herschlag, D. (1999) *RNA* **5,** 158–166.
43. Herschlag, D. & Khosla, M. (1994) *Biochemistry* **33,** 5291–5297.
44. Moran, S., Kierzek, R. & Turner, D. H. (1993) *Biochemistry* **32,** 5247–5256.
45. Michel, F. & Westhof, E. (1990) *J. Mol. Biol.* **216,** 585–610.
46. Shan, S., Narlikar, G. J. & Herschlag, D. (1999) *Biochemistry* **38,** 10976–10988.
47. Fersht, A. (1984) *Enzyme Structure and Mechanism* (Freeman, New York).
48. Szewczak, A. A., Ortoleva-Donnelly, L., Ryder, S. P., Moncoeur, E. & Strobel, S. A. (1998) *Nat. Struct. Biol.* **5,** 1037–1042.
49. Szewczak, A. A., Kosek, A. B., Piccirilli, J. A. & Strobel, S. A. (2002) *Biochemistry* **41,** 2516–2525.
50. Zhuang, X. W., Bartley, L. E., Babcock, H. P., Russell, R., Ha, T. J., Herschlag, D. & Chu, S. (2000) *Science* **288,** 2048–2051.
51. Bevilacqua, P. C., Sugimoto, N. & Turner, D. H. (1996) *Biochemistry* **35,** 648–658.
52. Porschke, D. & Eigen, M. (1971) *J. Mol. Biol.* **62,** 361–381.
53. Serra, M. J. & Turner, D. H. (1995) *Methods Enzymol.* **259,** 242–261.
54. Bartley, L. E., Zhuang, X. W., Das, R., Chu, S. & Herschlag, D. (2003) *J. Mol. Biol.*, in press.
55. Shan, S. & Herschlag, D. (2000) *RNA* **6,** 795–813.
56. Liao, X. M., Anjaneyulu, P. S. R., Curley, J. F., Hsu, M., Boehringer, M., Caruthers, M. H. & Piccirilli, J. A. (2001) *Biochemistry* **40,** 10911–10926.
57. Narlikar, G. J. & Herschlag, D. (1997) *Annu. Rev. Biochem.* **66,** 19–59.
58. Burkard, M. E., Turner, D. H. & Tinoco, I. (1999) in *The RNA World*, eds. Gesteland, R. F., Cech, T. R. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed., pp. 233–264.
59. Baldwin, R. L. & Rose, G. D. (1999) *Trends Biochem. Sci.* **24,** 26–33.
60. Gerstein, M., Lesk, A. M. & Chothia, C. (1994) *Biochemistry* **33,** 6739–6749.
61. Herschlag, D. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 6921–6925.
62. Suh, E. R. & Waring, R. B. (1990) *Mol. Cell. Biol.* **10,** 2960–2965.
63. Peluso, P., Herschlag, D., Nock, S., Freymann, D. M., Johnson, A. E. & Walter, P. (2000) *Science* **288,** 1640–1643.
64. Wu, L. C., Tuot, D. S., Lyons, D. S., Garcia, K. C. & Davis, M. M. (2002) *Nature* **418,** 552–556.
65. Rose, M. A. & Weeks, K. M. (2001) *Nat. Struct. Biol.* **8,** 515–520.
66. Webb, A. E., Rose, M. A., Westhof, E. & Weeks, K. M. (2001) *J. Mol. Biol.* **309,** 1087–1100.
67. Weeks, K. M. & Cech, T. R. (1996) *Science* **271,** 345–348.
68. Solem, A., Chatterjee, P. & Caprara, M. G. (2002) *RNA* **8,** 412–425.
69. Shan, S. & Herschlag, D. (2002) *RNA* **8,** 861–872.