

- McCall, M., Brown, T., & Kennard, O. (1985) *J. Mol. Biol.* 183, 385-396.
- McGhee, J. D., & Felsenfeld, G. (1980) *Annu. Rev. Biochem.* 49, 1115-1156.
- McGhee, J. D., & Felsenfeld, G. (1982) *J. Mol. Biol.* 158, 685-698.
- McGhee, J. D., Wood, W. I., Dolan, M., Engel, J. D., & Felsenfeld, G. (1981) *Cell* 27, 45-55.
- Nickol, J., Behe, M., & Felsenfeld, G. (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79, 1771-1775.
- Noll, M. (1974) *Nucleic Acids Res.* 1, 1573-1578.
- Ramsay, N., Felsenfeld, G., Rushton, B., & McGhee, J. D. (1984) *EMBO J.* 3, 2605-2611.
- Rhodes, D. (1979) *Nucleic Acids Res.* 6, 1805-1816.
- Ruiz-Carrillo, A., Jorcano, J. L., Eder, G., & Lurz, R. (1979) *Proc. Natl. Acad. Sci. U.S.A.* 76, 3284-3288.
- Satchwell, S. C., Drew, H. R., & Travers, A. A. (1986) *J. Mol. Biol.* 191, 659-675.
- Simon, R. H., & Felsenfeld, G. (1979) *Nucleic Acids Res.* 6, 689-696.
- Simpson, R. T., & Whitlock, J. P., Jr. (1976) *Cell* 9, 347-353.
- Simpson, R. T., & Kunzler, P. (1979) *Nucleic Acids Res.* 6, 1387-1415.
- Wu, H.-Y., & Behe, M. J. (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81, 7284-7287.
- Zhurkin, V. B. (1985) *J. Biomol. Struct. Dyn.* 2, 785-804.

Specificity of Arginine Binding by the *Tetrahymena* Intron[†]

Michael Yarus

Department of Molecular, Cellular, and Developmental Biology, Campus Box 347, University of Colorado, Boulder, Colorado 80309

Received July 19, 1988; Revised Manuscript Received October 4, 1988

ABSTRACT: L-Arginine competitively inhibits the reaction of GTP with the *Tetrahymena* ribosomal self-splicing intron. In order to define this RNA binding site for arginine, K_i 's have now been measured for numerous arginine-like competitive inhibitors. Detailed consideration of the K_i 's suggests a tripartite binding model. The dissociation constants of the inhibitors can be consistently interpreted if the guanidino group of arginine binds in the GTP site by utilizing the H-bonds otherwise made to the N1-H and 2-NH₂ of the guanine pyrimidine ring. The positive charge of the arginine guanidino group also enhances binding. A second requirement is for the precise length of the aliphatic arm connecting the guanidino with the α -carbon. The positive charge of the α -amino group is the third feature essential to effective inhibition. The negative carboxyl charge of arginine inhibits binding, and the substituents on the α -carbon are probably oriented, with the α -amino group near the phosphate backbone of the RNA. This orientation contributes strongly to the L stereoselectivity of the amino acid site on the RNA. When spaced optimally, net contribution to the free energy of binding is of the same order for the guanidino group and for the arginine α -carbon substituents, but the guanidino apparently contributes more to binding free energy. Taken together, these observations extend the previous binding model [Yarus, M. (1988) *Science (Washington, D.C.)* 240, 1751-1758]. The observed dependence of binding on universal characteristics of amino acids suggests that RNA binding sites with other amino acid specificities could exist.

Among the 20 biological amino acids, L-arginine alone is an effective inhibitor of the group I self-splicing reaction carried out by the *Tetrahymena* ribosomal intron (Yarus, 1988). L-Arginine inhibits more strongly than D-arginine, and inhibition is competitive with the attack of GTP on the RNA precursor. Arginine appears to block the guanosine (rG) site, without further effects on catalysis. Arginine could bind in the rG site, using the structural similarity between the guanidino group of the arginine side chain and an isosteric set of atoms on the guanosine six-membered ring (Yarus, 1988).

Such specific binding of an amino acid by an RNA has not been previously studied. The nature of the binding interaction bears on the origin of the genetic code, on the possibility of amino acid substrates for RNA enzymes, and on potential regulatory interactions between RNAs and amino acids (Yarus, 1988). Therefore, the arginine binding site on the intron has been probed with a set of arginine analogues which exemplify or alter all parts of the arginine molecule. The thermodynamic dissociation constants of these materials (the K_i 's) are presented here. These data distinguish two possible

binding modes for arginine and further resolve the sources of the binding free energy and L stereoselectivity shown by the RNA.

EXPERIMENTAL PROCEDURES

Inhibitors. Materials of the highest commercially available purity were used, after storage in vacuo over P₂O₅. Solutions were adjusted to pH 8 if addition of inhibitor to splicing reactions at maximal concentrations changed the pH > 0.1.

L- and D-arginine from Chemical Dynamics, Fluka, and Sigma were compared without detectable differences. Guanidinoformic acid, guanidinoacetic acid, methylguanidine hydrochloride, ethylguanidine hydrochloride, butylguanidine hydrochloride, N^α-acetyl-L-arginine, N^α-benzoyl-L-arginyl ethyl ester, 6-guanidinocaproic acid, and 5-guanidinovaleric acid were obtained from Aldrich or the Bader Library of Rare Chemicals. L-2-Amino-3-guanidinopropionic acid and L-2-amino-4-guanidinobutyric acid were from Chemical Dynamics. Guanylylurea sulfate was obtained from Eastman Kodak. Agmatine, L-argininamide, L-arginine methyl ester, L-arginine ethyl ester, L-leucyl-L-arginine, L-argininyl-L-leucine, L-homoarginine, L-canavanine, L-citrulline, creatine, creatinine,

[†] This work was supported by NIH Research Grant R37 GM30881.

L-ornithine, 3-guanidinopropionic acid, 4-guanidinobutyric acid, guanidosuccinic acid, urea, L-2-hydroxy-5-guanidinovaleric acid, and L-lysine were obtained from Sigma. Guanidine hydrochloride was obtained from Schwartz BioResearch.

Precursor RNA. Plasmid TT1A3-T7 contains the 482-nt *ThaI* fragment which spans the intron of the *Tetrahymena* rDNA in pIVS11 (Kruger et al., 1982). This *ThaI* fragment was terminated with *HindIII* linkers and inserted into the *HindIII* site of pT7-2 (U.S. Biochemical Corp.) to construct pTT1A3-T7. The splicing precursor used for these measurements is a T7 RNA polymerase runoff transcript of pTT1A3-T7 cut at its unique *EcoRI* site, which is adjacent to the *HindIII* site in the pT7-2 polylinker. The result is a precursor RNA containing a 5' exon of 42 nt, the 413-nt intron, and an 82-nt 3' exon.

Transcription. Two micrograms of *EcoRI*-cut pTT1A3-T7 was transcribed in a total volume of 100 μ L containing 40 mM Tris buffer, pH 8, 1 mM spermidine, 0.01% Triton X-100, 6 mM $MgCl_2$, 5 mM DTT, and 1 mM each of the four nucleoside triphosphates, unlabeled. About 1000 units of T7 RNA polymerase gave ca. 25% conversion of the triphosphates to RNA in 1 h at 30 °C, with no visible production of splicing products on subsequent gel radioautograms (if labeled triphosphates were used to visualize them). Transcription was stopped by heating to 65 °C for 5 min after addition of 50 μ L of saturated urea (room temperature), 50 mM EDTA, and 0.05% xylene cyanol and bromophenol blue. Precursor RNA was resolved from residual nucleotides on Sephadex G-50 fine in 10 mM Tris, pH 8, 1 mM EDTA, and 500 mM NaCl. The excluded peak from a 9-mL column bed (the 1.5 mL which follows 3.0 mL of eluant) was precipitated with ethanol, washed, and stored at -20 °C in 10 mM Tris, pH 8, and 1 mM EDTA.

Splicing. The reaction mixture used consisted of 5 μ L of 50 mM HEPES, 30 mM NaCl, 10 mM $MgCl_2$, and 0.2 mM EDTA, pH 8.0. Incubations were 2–5 min at 37 °C, chosen to keep production of the IVS-3' exon in the linear (initial velocity) region of the kinetics. It is important to note that precursor RNA stored frozen must be reactivated (2.5 min at 65 °C in 1 mM free Mg^{2+}) for reproducible maximum splicing velocities. Reactions were stopped by addition of 2.5 μ L of the urea, EDTA, and dye solution listed under *Transcription*. Unlabeled precursor transcript was ca. 0.1–0.3 μ M chains (attributing all A_{260} to active molecules), and [α - ^{32}P]GTP was present at 0.033 μ M (3000 C/mmol). These conditions measure initial velocity of splicing (see below) and consume only a small fraction of the GTP and precursor RNA. Both GTP and precursor RNA are therefore present at essentially their initial concentrations throughout the splicing reaction. The free GTP concentration is also much smaller than K_m (Yarus, 1988), which is important to the analysis of the results (below).

For K_i for L-arginine was redetermined during this work (data in Figure 2), and this value is in reasonable agreement with the previous independent, more extensive experiment (K_i = 4.0 mM; Yarus, 1988). All data for D-arginine, L-arginine methyl and ethyl esters, L-argininamide, agmatine, L-arginyl-L-leucine, L-leucyl-L-arginine, L-2-hydroxy-5-guanidinovaleric acid, guanidinoacetic acid, guanidine hydrochloride, and L-lysine are taken from Yarus (1988).

Analysis. Splicing products end labeled with [^{32}P]GTP were resolved by gel electrophoresis conducted in TBE plus 8 M urea, in vertical slabs of 5% acrylamide (20:1 bis). TBE is 0.1 M Tris, 0.1 M boric acid, and 2 mM EDTA, pH 8.3. Gels were dried onto DEAE paper to retain all radioactivity. Dried

gels were radioautographed and then sliced for Cerenkov scintillation counting, the film being used as a template. In these experiments, the radioactivity in the IVS-3' exon band was normalized to that from an uninhibited reaction done in triplicate for each gel. The result is f , the fraction of the control splicing velocity.

Determination of K_i utilizes

$$f^{-1} = 1 + [K_m/(K_m + S)][I/K_i] \quad (1)$$

f is the fraction of the rate of reaction of an uninhibited control, K_m is the Michaelis constant for the substrate, S is substrate concentration, K_i is the dissociation constant of the inhibitor, and I is inhibitor concentration. This is a rearrangement of the standard equation for competitive inhibition of an enzyme. In these experiments $K_m/(K_m + S)$ is effectively 1 because low GTP concentrations ($K_m \gg [GTP]$) are used. Under these conditions

$$f^{-1} = 1 + I/K_i \quad (2)$$

and plots of f^{-1} versus I are linear and have a slope which is the binding constant of the inhibitor, or the reciprocal of the dissociation constant, K_i .

Errors. These plots were fitted by least-squares methods, and the K_i 's are given in millimolar, with a standard error (SE) propagated to the reciprocal of the slope of the line under the assumption that the variance of all measured fractional velocities is the same.

The gel assay for splicing velocity is of moderate precision for single assays but is subject to occasional excursions because of the many operations required to produce each value. Despite these qualities, evident in Figure 2, this method gives good precision for the dissociation constant of any inhibitor which can be used at its K_i or higher.

However, inhibitor concentrations were restricted to 30–60 mM in order to avoid complications which might result from gross changes in the ionic composition of the reactions. Creatinine, which is nonionic, is an exception and was used at ≤ 200 mM. Weaker inhibition (large K_i 's) is less accurately determined, because inhibition approaches the limit of accuracy of the assay. Weak inhibitions are also more easily perturbed by nonspecific effects. Comparisons of the K_i 's of the weaker inhibitors must therefore respect the standard errors, which tend to reflect the considerable uncertainty of such measurements.

Very weak inhibitors are sometimes characterized in the tables as $K_i > 200$ –250 mM. Different limits cited are due to differences in the largest concentration that was tried without significant splicing inhibition.

RESULTS

Inhibition Is Competitive. Estimation of the K_i by the method used here yields valid dissociation constants if inhibition is either competitive or alternatively noncompetitive, with complete splicing inhibition if the inhibitor is bound. To determine the type of inhibition for this variety of inhibitors, splicing inhibition was measured at 0.033, 2.0, and 11.2 μ M GTP for a constant concentration of L-arginine (at 3.1 times the apparent K_i), D-arginine (2.4 K_i), L-argininamide (6.8 K_i), arginylleucine (2.1 K_i), agmatine (4.6 K_i), arginine methyl ester (3.2 K_i), arginine ethyl ester (3.6 K_i), L-homoarginine (1.1 K_i), guanidinoacetic acid (0.63 K_i), guanylurea (2.1 K_i), L-lysine (0.65 K_i), L-2-amino-3-guanidinopropionic acid (0.88 K_i), and L-2-hydroxy-5-guanidinovaleric acid (4.6 K_i).

For all inhibitors tested, apparent inhibition decreased (see eq 1) as GTP concentration increased. Decreased inhibition was quantitatively consistent with inhibition competitive with

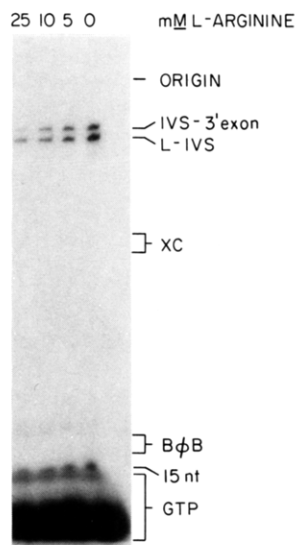


FIGURE 1: Inhibition of splicing by L-arginine. XC, position of xylene cyanole dye; B ϕ B, position of bromophenol blue; IVS-3' exon, the two-thirds molecule created by the insertion of G at the 5' exon intron junction; L-IVS, linear intron, the immediate product of joining of the exons; 15nt, 5' pentadecanucleotide released when L-IVS circularizes (Zaug et al., 1983).

GTP ($K_m = 2.1 \pm 0.4 \mu\text{M}$ under these conditions; Yarus, 1988). Thus, increased GTP excludes (competes with) all of this set of arginine analogues, which includes all the strong splicing inhibitors analyzed here. Previous analysis showed (Yarus, 1988) that other catalytic activities (specific hydrolysis and circularization) of the intron were not inhibited by L-arginine, consistent with simple competition for the GTP site.

Dissociation constants (K_i 's) for these inhibitors have been measured by determination of the initial rate of production of the first splicing intermediate, IVS-3' exon (Zaug et al., 1983). The relative rate of reaction of GTP with a T7 transcript of pTT1A3-T7 is measured in the presence of a series of inhibitor concentrations. Figure 1 shows a radioautogram of the three G-labeled RNA splicing products in the presence of 0, 5, 10, and 25 mM L-arginine. Progressive inhibition by arginine is evident from the decreases in the radioactivity in the gel zones of all the splicing intermediates. The reciprocal fraction of the rate, when plotted versus the inhibitor concentration, gives a line whose slope is proportional to the association constant of the inhibitor (proportional to the inverse of the dissociation constant: eq 2). Such data, for inhibitors which illustrate the range of K_i 's in this study, are shown in Figure 2.

Least-squares lines are fitted to these data, and the resulting K_i and its standard error in millimolar are presented in later figures. Some of the data presented, derived by the same procedure, is taken from previous work (Yarus, 1988).

Data for the Derivatives. Figure 3 shows measured K_i 's for a set of analogues in which variants of an arginine-like side chain are characterized. These data suggest the specificity of the RNA interaction with the distal part of the arginine side chain.

Figure 4 shows the binding measured for compounds related to the guanidinium ion. These data were collected to evaluate the sources of the binding free energy between the RNA and the guanidino group of arginine.

Figure 5 shows K_i 's for a set of derivatives which all have a guanidino and carboxyl group, but vary in the length of the aliphatic arm which connects them. In some cases, members of this series are matched by a compound which also possesses

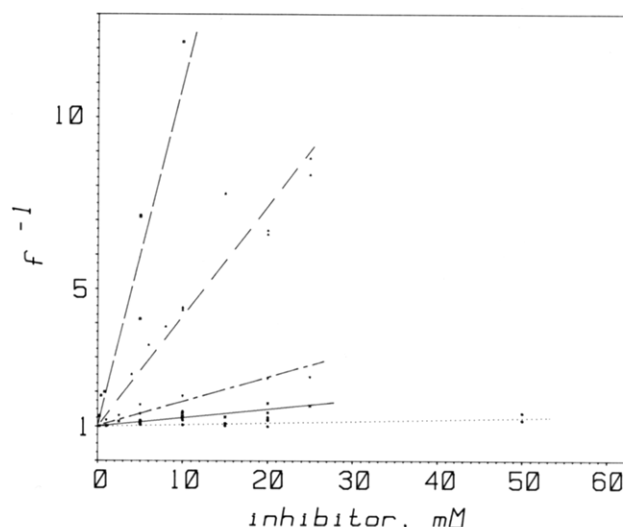


FIGURE 2: Plot of the reciprocal of the relative velocity of GTP reaction (f^{-1}) versus the millimolar concentration of inhibitor. The slope of the line is proportional to the binding constant (see Experimental Procedures). Squares, upper dashed line, inhibition by agmatine; triangles, lower dashed line, L-arginine; inverted triangles, dashed and dotted line, L-homoarginine; X, solid line, L-lysine; circles, dotted line, guanidine hydrochloride.

an α -amino group (an L-amino acid). The affinities of the RNA for this set of compounds illuminate the role of the overall size of the arginine analogue and, in addition, clarify the specific contribution of the α -amino group.

Finally, Figure 6 shows the association of a set of derivatives of the α -NH₂ and carboxyl groups of arginine, as well as the previously measured (Yarus, 1988) stereoselectivity for L-arginine. These measurements resolve the roles of the amino and carboxyl groups in binding and stereoselectivity.

DISCUSSION

Explicit structural interpretation of these binding data is based on the model previously presented (Yarus, 1988) and shown again in Figure 7A. Arginine is assumed to be bound to RNA in its most stable conformation (Ponder & Richards, 1987), superposed upon the position otherwise occupied by the normal splicing substrate, rG or GTP. The location of the amino acid is determined by the hypothesis that the guanidino group forms the same set of hydrogen bonds as the guanine N1-H and 2-NH₂ (Yarus, 1988; Bass & Cech, 1984). Note however that the arginine side chain has four rotatable bonds and therefore many conformations which do not differ greatly in stability. Because the RNA binding site also may be flexible, the interpretation below, usually conducted in terms of a single conformation for both RNA and amino acid analogue, should be considered a useful approximation.

Uniqueness of the Guanidino Group. Better binding by L-arginine than by L-lysine and other L amino acids (Yarus, 1988) suggests that guanidino hydrogen bonding is an important contributor to arginine binding. Figure 3 presents arginine, drawn so that the resemblance to guanosine is evident. The 1, 2, and 6 atoms of the guanine pyrimidine ring and the 2' and 3' atoms of the ribose are indicated for reference.

L-Canavanine is a metabolic poison which distorts proteins because it is incorporated into tRNA by the arginyl-tRNA synthetase (Allende & Allende, 1964). Rat liver arginyl-tRNA synthetase makes a 7-fold distinction against canavanine ($K_i > K_m$ for arginine; Allende & Allende, 1964). The RNA site, notably, selects more strongly than rat liver arginyl-tRNA synthetase against the substitution of -O- for -CH₂- adjacent to the guanidino group, making a 30-fold distinction between

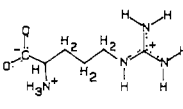
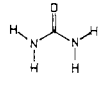
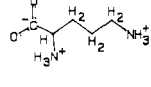
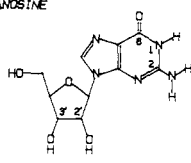
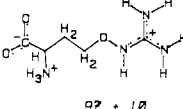
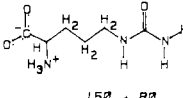
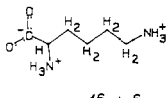
NAME	L-ARGININE	UREA	L-ORNITHINE
STRUCTURE			
$K_i \pm SE, \text{mM}$	3.2 ± 0.3	1200	1200
GUANOSINE	L-CANAVANINE	L-CITRULLINE	L-LYSINE
			
	97 ± 10	150 ± 80	46 ± 6

FIGURE 3: Splicing inhibition by derivatives of L-arginine. Guanosine with its 2', 3', 1, 2, and 6 atoms marked is shown for reference.

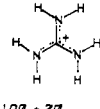
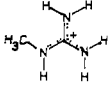
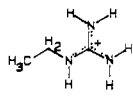
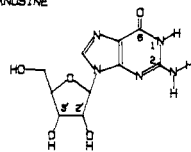
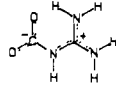
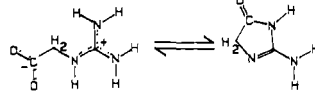
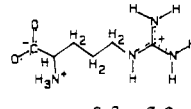
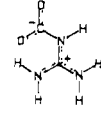
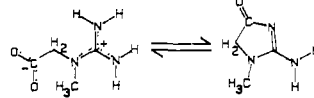
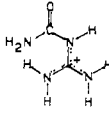
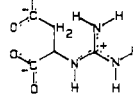
NAME	STRUCTURE	$K_i \pm SE, \text{mM}$
GUANIDINIUM		100 ± 70
METHYL GUANIDINIUM		140 ± 60
ETHYL GUANIDINIUM		140 ± 20
GUANOSINE		
GUANIDINO FORMIC ACID		70 ± 0
2-GUANIDINO ACETIC ACID \rightleftharpoons GLYCODYLAMIDINE		50 ± 10
L-ARGININE		3.2 ± 0.3
GUANIDINO FORMIC ACID		70 ± 0
CREATINE \rightleftharpoons CREATININE		240 ± 70 400 ± 140
GUANYL UREA		11 ± 1
2-GUANIDINO SUCCHINIC ACID		1200

FIGURE 4: Splicing inhibition by compounds which resemble the distal side chain of arginine.

L-arginine and L-canavanine (Figure 3). This result supports the significance of the distal arginine side chain and demonstrates that the RNA site is capable of subtle distinctions. In this case, the added $-\text{O}-$ on the amino acid is adjacent to a hydrogen which may be donating a hydrogen bond (Bass & Cech, 1984; compare Figure 7), and the oxygen may clash with the hydrogen-bond acceptor on the RNA, perhaps another oxygen.

Imitation of other aspects of the hydrogen-bonding pattern of guanine does not yield effective inhibitors of splicing. Urea does not inhibit detectably, though L-citrulline is an efficient inhibitor, 47-fold weaker than arginine (Figure 3). Both emulate the hydrogen-bonding pattern of the $\text{C6}=\text{O}$ and $\text{N1}-\text{H}$ of guanine, and citrulline is the same size and configuration as arginine, save for the ureido rather than guanidino group of the side chain. Given the likely assumption that the electrostatic potential is negative throughout the arginine site (data below), the difference between L-citrulline and L-arginine may be due in part to the positive charge of the guanidino group, in part to an extra H-bond to the guanidino group [compare Bass and Cech (1984)].

It is likely that the spacing between the groups on the α -carbon and the positive charge of the side chain is an influential variable. L-Lysine presents positive charges at the same spacing as in L-arginine, though it does not emulate the guanine hydrogen-bonding pattern. L-Lysine binding is easily detectable, though 14-fold weaker than that of arginine. Because L-ornithine (Figure 3), an analogue of L-lysine differing only in the length of the side chain, does not inhibit detectably, the longer side chain evidently is required. The

importance of these charges and their spacing is confirmed more directly below (Figures 4–6).

Analogues of the Distal Side Chain. Figure 4 presents an effort to further resolve the role of the side chain by use of inhibitors which resemble fragments of the distal arginine side chain.

Guanidinium has a potential entropic binding advantage over arginine because it is symmetrical and may bind in three orientations, yet guanidinium (Figure 4) is a poor inhibitor of splicing, 59-fold weaker than arginine. The addition of aliphatic elements to guanidinium stimulates binding slightly, if at all (methyl- and ethylguanidinium, Figure 4). If the one- and two-carbon additions are not aliphatic, but instead contain the polar carboxyl group, binding is apparently somewhat improved (guanidinoformic acid and guanidinoacetic acid, Figure 4).

This carboxyl effect is unexpected; it is the only case in which addition of a negative charge to an inhibitor apparently augments the free energy of binding to the negatively charged RNA. I propose that special factors more than compensate the negative charges on guanidinoformic acid and guanidinoacetic acid. One possible compensating factor is shown in the figure by redrawing of guanidinoformic acid to make clear that it potentially mimics all three hydrogen-bonding groups of the guanosine pyrimidine ring (rightward in the figure). This mimicry is possible despite strong steric hindrance between the carboxyl and the guanidino $-\text{NH}_2$, which forces these groups to be somewhat nonplanar, and therefore an imperfect replica of guanine (not shown).

A further analogue suggests that the electrostatic potential

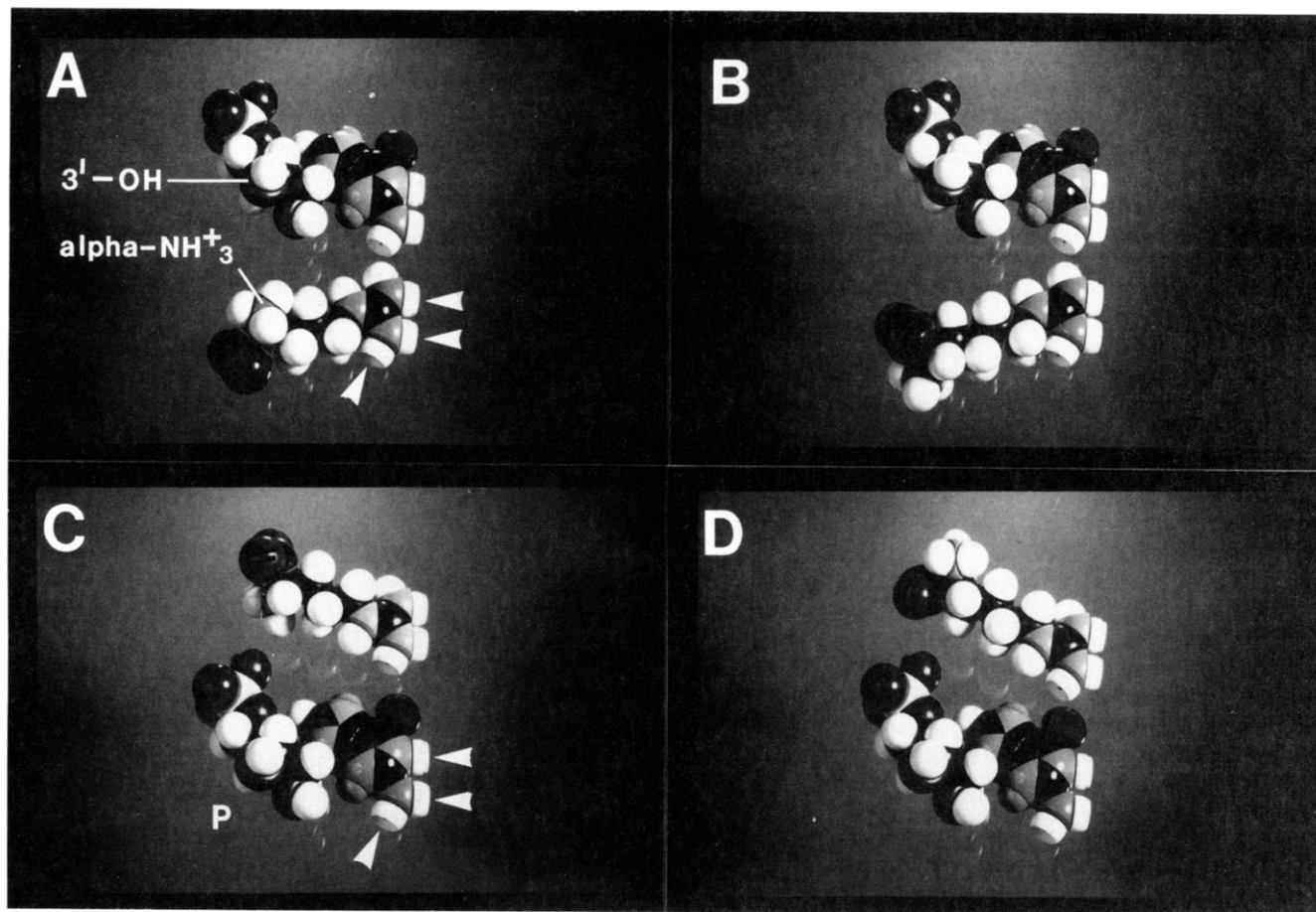


FIGURE 7: A comparison of arginine analogues and GMP. Space-filling models of the two are shown side-by-side, as if the guanidino group of the amino acid were oriented by binding to the same groups as the analogous atoms in the guanine pyrimidine ring. Relative positions of the two may be envisioned by mentally sliding the amino acid to superposition on the GMP. (Panel A) Binding mode I, L-arginine below GMP. White arrowheads indicate hydrogen bonds which may also be formed to guanine (Figure 7C). (Panel B) Binding mode I, L-homoarginine below GMP. (Panel C) Binding mode II, L-arginine above GMP. P suggests the position of the 5' exon-intron junction phosphodiester; white arrowheads indicate hydrogen bonds which may also be formed to the guanidino of arginine (panel A). (Panel D) Binding mode II, L-homoarginine above GMP.

augment the net free energy of binding observed for ethylguanidinium or guanidinoacetic acid. In the next section, these additional interactions are identified.

Role of the Aliphatic Arm and the α -Amino Group. In Figure 5, the series of guanidino acids is carried out to 6-guanidinocaproic acid, and the largest four derivatives are compared to their related L amino acids.

First, though the stimulation varies with chain length, the addition of an α -amino group stimulates binding in every pair of compounds. Such a general stimulation without regard to position is likely to be electrostatic, attributable to the interaction of the positive α -amino group with a negative potential from the RNAs phosphate backbone.

Second, the effect of the amino group becomes particularly marked for chains the length of arginine. Strong stimulation by the α -amino group persists for homoarginine. That is, as the aliphatic side chain is lengthened, the α -amino group probably approaches a backbone phosphate situated near the position of the α -amino of bound arginine.

Third, extension of the aliphatic portion of the side chain of these compounds beyond that of arginine reduces the strength of binding for both carboxyl and α -amino derivatives. This result suggests that the size of the arginine/rG site is such that it can be precisely spanned by a five-carbon backbone attached to the guanidino group. Longer chains can be bound more weakly, perhaps because they fold the aliphatic side chain by assuming a somewhat more compact, but less stable, rotamer.

One consequence of these observations is that the rG site on the intron is surprisingly well formed for specific arginine binding. The bound ligand must be an amino acid whose α -amino is spaced correctly from the guanidino group. Arginine is precisely the optimal size, and addition or subtraction of one methylene group decreases binding free energy. The effects of side-chain length may also include rotation of the α -carbon, which is oriented differently in odd- and even-length chains (compare arginine and homoarginine in Figure 7).

Two Possible Modes of Binding. The data of Figure 5 also clarify a potential ambiguity in the resemblance of guanine and guanidino. Figure 7 compares space-filling models of arginine and rG. As shown in the figure, there are two binding moles for arginine, related by 180° rotation around an axis of the guanidino group. Both modes were mentioned, and mode I was pictured in the previously published discussion (Figure 7A,B). Mode II (Figure 7C,D) forces a different orientation of the extended arginine side chain, roughly along the long axis of the purine ring instead of along the long axis of the rG molecule as in mode I.

As can be appreciated by inspection of the figure, the data in Figure 6 argue for mode I. With the relative orientation of mode I, arginine has roughly the same dimensions as rG in the 3'-endo anti conformation (Figure 7A). However, homoarginine (and 6-guanidinocaproic acid) with its guanidino group engaged would extend beyond the edge of the space required for rG (Figure 7B). This steric situation corresponds well to the measured dissociation constants (Figure 6). In

contrast, binding via mode II (Figure 7C,D) would not distinguish arginine and homoarginine (or 5-guanidinovaleric acid and 6-guanidinocaproic acid). Mode II requires that both extend well beyond the volume defined by the guanine ring of rG.

Assuming mode I, the phosphate approached by the α -amino group of arginine can be tentatively identified as that involved in the exon-intron junction, UpA in *Tetrahymena* (Zaug & Cech, 1982). This is so because of the near superposition of the α -amino of arginine and the 3'-OH of the rG ribose (Figure 7A), which must attack that same phosphodiester bond to initiate splicing. Assuming that the positions of the groups in the rG site are not substantially changed when arginine occupies the same site, then the negative charge of that phosphate would stabilize binding of L-arginine, as observed. By the same token, this phosphate would play a part in the selection against D-arginine (Yarus, 1988), which would have to assume a less stable rotamer to bring its α -amino close to the phosphate.

Derivatives of the Arginine α -Amino and Carboxyl. Figure 6 shows the dissociation constants for arginine congeners with modified carboxyl groups (left) and α -amino groups (right). As previously argued (Yarus, 1988), it is likely that the amino group is more hindered than the carboxyl. For example, compare the results of addition of leucine to the two groups in the figure. Less specifically, note that dissociation constants are much greater for the amino derivatives (right) than for the carboxyl derivatives (left).

However, the disparity also reflects the contribution to binding made by the α -amino charge, as revealed by the data of Figure 5. In fact, the data in Figures 5 and 6 taken together show that the groups on the α -carbon are probably oriented in the field of the nearby backbone phosphate(s), α -amino toward the backbone, carboxyl away.

This orientation follows because the maximal effect of the carboxyl negative charge is 3–5-fold: agmatine (decarboxylated arginine) is approximately 3-fold better than arginine, and argininamide is less than 5-fold better, Figure 6. This is an upper limit on the effect of the charge because it includes the steric effect of the carboxyl in the case of agmatine and includes possible hydrogen bonding by the amide in the case of argininamide. Note that when the carboxyl is enlarged and the charge is removed by nonpolar substitutions (arginine methyl and ethyl esters, Figure 6), binding may be stimulated, but the factor is smaller and perhaps insignificant for these larger substituents.

In contrast, the minimum contribution of the α -amino charge is 25-fold: compare 5-guanidinovaleric acid and arginine in Figure 5. This is a minimum because it includes the potentially inhibitory effect of addition of the amino group's volume. Therefore, the likely electrostatic effect is substantially stronger for the amino group, placing it nearer the phosphate.

The sole case in which the α -amino is removed, and binding improved, is L-2-hydroxy-5-guanidinovaleric acid (Figure 6). This exceptional result can also be accommodated by the binding model. In this case, the replacement of the α -amino by an α -hydroxyl makes possible a strong hydrogen bond, perhaps to one of the oxygens of the nearby UpA intron junction phosphate. By hypothesis, the H-bond slightly more than overcomes the effect of the loss of the positive charge of the α -NH₂. The simplest binding model therefore suggests that the phosphate is within atomic dimensions of, or easily brought to, the position of the α -hydroxyl group. This is again consistent with binding in mode I, which would place the

α -hydroxyl and the ribose 3'-OH at similar positions (compare Figure 7A), so that the bond to the ribose 3'-OH deduced by Bass and Cech (1986) can plausibly also be formed to the α -hydroxyl group of the amino acid derivative.

A Summary of the Model of the Binding Site. The most remarkable result of this analysis is that a single notion about similarity between arginine and rG can be elaborated into a self-consistent explanation for everything that is known of splicing inhibition by arginine analogues. This consistency supports the conclusion, from kinetic data, that splicing inhibition is predominantly or entirely due to a single bound arginine, acting within the intron's unique site for GTP. In addition, both the binding free energy and the stereoselectivity for the amino acid are putative consequences of groups which must be in identical positions (Figure 7A,C) to account for the hydrogen bonding and reactivity of the normal splicing substrate, guanosine. The tripartite binding model below is therefore very economical of hypotheses.

By reference to Figure 7A, the three essential aspects of the arginine binding site suggested by these experiments may be envisioned. At the right, a system of hydrogen bonds (white arrowheads), which may be formed both to arginine (Figure 7A) and to guanine (Figure 7C), defines the position of the amino acid and contributes to binding free energy. It is also likely that the positive charge of the guanidino group facilitates binding (Figure 3).

On the left, the most stable position for the α -amino and carboxyl groups is determined by the orienting effect of the negative charge and volume of groups attached to the position marked P. P represents the 5' exon-intron junction phosphodiester, ...UpA..., and should be visualized at the level of the 3'-OH in Figure 7C. The hindrance encountered by 6-guanidinocaproic acid and L-homoarginine (Figure 5) is one experimentally detectable expression of the presence of the exon-intron junction at P (Figure 7A,C). The junction phosphodiester can be visualized according to the suggestion of Cedergren et al. (1987). It would be centered at P (or right and downward from the arginine α -amino in Figure 7A), oriented so that attack of the 3'-OH of rG takes place outward and opposite the 3'O-P bond of the junction phosphodiester, UpA. The steric effect of UpA and, probably even more importantly, the electrostatic effects of the phosphate of UpA and the nearby backbone produce the measured stereoselection for L-versus D-arginine.

This description extends that originally presented (Yarus, 1988). At the time the binding model was originally formulated, only the steric asymmetry of the arginine site was clear from the derivatives whose dissociation constants were known. Thus, the contribution to binding and stereoselection of the oriented α -NH₂ and -COO⁻ in the electrostatic field of the RNA backbone has only become obvious in this work (Figures 5 and 6).

Suppose that a binding site like this one accounts for the original association of amino acid and RNA sequences which became the genetic code. If the spatial relation between the intron junction phosphodiester and the hydrogen-bonding groups for rG is maintained by a stable and common RNA structure, rather than arising as a result of some idiosyncrasy of the *Tetrahymena* intron, it would be possible that this common RNA structure could have determined the evolutionary choice of L rather than D amino acids in proteins (Yarus, 1988).

The third element of the binding model is the requirement that the guanidino group be connected to the arginine α -carbon by a side chain precisely the length of that in arginine. One

methylene group shorter (L-2-amino-4-guanidinobutyric acid, Figure 5) or one longer (L-homoarginine, Figure 5) produces binding distinctly inferior to that of L-arginine itself. The characteristics of the space bridged by the side chain are suggested by the K_i 's of the carboxylic acids on the left in Figure 5. In the absence of the strong tendency to approach the UpA contributed by the amino group, these dissociation constants vary only 3-fold, until the longest member of the series apparently becomes too large to fit the site. Thus, the binding site between its polar H-bonding face and the polar UpA contact appears to be homogeneous and chemically similar to water. That is, the extending side chains are neither strongly stabilized nor destabilized relative to their solution states because binding free energy is almost constant. For example, the hydrophobic part of the side chain does not encounter a strong hydrophobic contact which would stimulate binding, nor is the carboxyl forced to approach the phosphate backbone, which would hinder binding. These results, however, may be tied to the particular probe employed. The guanidino carboxylic acids are both more flexible and smaller than the guanosine that occupies this same space during space initiation. However, for the purpose of amino acid binding, the role of the methylenic part of the arginine side chain is well approximated as that of a featureless bridge.

Free Energies for Binding. This series of derivatives allows approximation of the specific contributions of the groups at each end of the L-arginine side chain. Guanidino hydrogen bonding contributes approximately the difference between L-lysine and L-arginine, 14-fold in the equilibrium constant or ca. 1.6 kcal/mol (Figure 3). The electrostatic effect of the guanidino group must be estimated more indirectly, but it increases the binding constant at least 4-fold (>0.8 kcal/mol; lysine vs ornithine, Figure 3). Another estimate of stimulation by the positive guanidino charge comes from the 6.4-fold (1.1 kcal/mol) effect of the charge on guanidinoformic acid (vs guanylyurea, Figure 4), which is near the same point in the site as the guanidino charge.

At the other end of the side chain, the α -amino stimulates binding 25-fold (1.9 kcal/mol; 5-guanidinovaleric acid vs arginine, Figure 5). This stimulation is partially compensated by the negative carboxyl 3-fold depression (0.6 kcal/mol, agmatine vs arginine, Figure 6). Thus the α -carbon substituents contribute ca. 1.3 kcal/mol and the guanidino ca. 2.6 kcal/mol. Therefore, the side-chain contribution to binding, almost entirely from the guanidino rather than methylenic groups, is larger than the contribution of the α -carbon substituents. This pattern explains the notable preference of this RNA site for L-arginine among the amino acids (Yarus, 1988).

Other Amino Acid Binding Sites. The smaller contribution of the α -carbon substituents in the binding model also defines the requirements for other potential RNA amino acid sites with different specificities. The α -carbon substituents are universal; thus, the addition of a side chain that can add a few kilocalories per mole, equivalent to at least two strong hydrogen bonds to an RNA, is required to produce binding with a biochemically significant free energy. Obvious among the possibilities are glutamine and asparagine, whose side chains would emulate the C6=O and N1-H of guanine itself (compare the discussion of L-citrulline above). Clearly other RNAs should be carefully tested for binding of these three amino acids.

Dissection of the Binding Site. This work shows that small fragments of the guanosine or arginine molecules are detectably bound by the RNA site. Guanylyurea (Figure 4) is the most striking example because it has a binding constant

of the same order as arginine. Guanylyurea may be usable as a probe for that part of the intron which hydrogen bonds to the pyrimidine ring of G. Such a probe would not necessarily be affected by most of the normal stacking interactions to G nor by interactions which normally are made to ribose or ribose phosphates. Competitive inhibitors which represent fragments of G and arginine may therefore allow dissection of the contribution of individual nucleotides to the rG site by measurement of K_i , using different intron sequences derived from site-directed changes or natural variation.

An Amino Acid Derivative as Substrate. As Figure 7A makes clear, it is conceivable that an amino acid derivative might participate in splicing. This would be of great interest as a model for the hypothetical prebiotic catalyst which joined amino-acid-like molecules and RNAs (Yarus, 1988; Weiner & Maizels, 1987). In particular, L-2-hydroxy-5-guanidinovaleric acid, which binds well (Figure 6), would be expected to place its α -hydroxyl appropriately for such a reaction. However, this compound does not participate in the splicing reaction at a detectable rate (M. Yarus, unpublished results). However, this may not be surprising, since L-2-hydroxy-5-guanidinovaleric acid, with its single hydroxyl, may be considered an analogue of deoxyguanosine, which also competitively inhibits splicing without participating in the reaction (Bass & Cech, 1986). There is therefore a possibility that a vicinal diol with other chemical characteristics like arginine would emulate rG and react covalently with the RNA precursor.

ACKNOWLEDGMENTS

Thanks to Art Zaug for his generosity in supplying pTT1A3-T7 and purified T7 RNA polymerase. Eric Christian, Jim Curran, Brian Hicke, Leslie Mounkes, Dennis Schultz, and Drew Smith made useful comments on a draft manuscript.

Registry No. GTP, 86-01-1; L-arginine, 74-79-3; L-canavanine, 543-38-4; guanidine, 113-00-8; urea, 57-13-6; L-citrulline, 372-75-8; L-ornithine, 70-26-8; L-lysine, 56-87-1; guanidinium, 25215-10-5; methylguanidinium, 64448-81-3; ethylguanidinium, 117941-32-9; guanidinoformic acid, 4390-68-5; 2-guanidinoacetic acid, 352-97-6; glycohydrazine, 503-86-6; creatine, 57-00-1; creatinine, 60-27-5; guanylyurea, 141-83-3; 2-guanidinosuccinic acid, 6133-30-8; 3-guanidinopropionic acid, 353-09-3; 4-guanidinobutyric acid, 463-00-3; 5-guanidinovaleric acid, 462-93-1; 6-guanidinocaproic acid, 6659-35-4; L-2-amino-3-guanidinopropionic acid, 14191-91-4; L-2-amino-4-guanidinobutyric acid, 14191-90-3; L-homoarginine, 156-86-5; L-arginyl-L-leucine, 1188-24-5; L-argininamide, 2788-83-2; L-arginine ethyl ester, 28696-31-3; L-arginine methyl ester, 2577-94-8; agmatine, 306-60-5; D-arginine, 157-06-2; L-leucyl-L-arginine, 26607-15-8; L-2-hydroxy-5-guanidinovaleric acid, 157-07-3; N-benzoyl-L-arginine ethyl ester, 971-21-1; N-acetyl-L-arginine, 155-84-0.

REFERENCES

- Allende, C. C., & Allende, J. E. (1964) *J. Biol. Chem.* 239, 1102-1106.
- Bass, B. L., & Cech, T. R. (1984) *Nature (London)* 308, 820-826.
- Bass, B. L., & Cech, T. R. (1986) *Biochemistry* 25, 4473-4477.
- Cannan, R. K., & Shore, A. (1928) *Biochem. J.* 22, 920-929.
- Cederger, R., Lang, B. F., & Gravel, D. (1987) *FEBS Lett.* 226, 63-66.
- Edgar, G., & Shiver, H. E. (1925) *J. Am. Chem. Soc.* 47, 1179-1188.
- Kenyon, G. L., & Rowley, G. L. (1971) *J. Am. Chem. Soc.* 93, 5552-5560.
- Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J.,

- Gottschling, D. E., & Cech, T. R. (1982) *Cell* (Cambridge, Mass). 31, 147-157.
- Ponder, J. W., & Richards, F. M. (1987) *J. Mol. Biol.* 193, 775-791.
- Weiner, A., & Maizels, N. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 7383-7388.
- Yarus, M. (1988) *Science* (Washington, D.C.) 240, 1751-1758.
- Zaug, A. J., & Cech, T. R. (1982) *Nucleic Acids Res.* 10, 2823-2838.
- Zaug, A. J., Grabowski, P. J., & Cech, T. R. (1983) *Nature* (London) 301, 578-583.

Exonucleolytic Proofreading by a Mammalian DNA Polymerase γ^{\dagger}

Thomas A. Kunkel^{*†} and Dale W. Mosbaugh[§]

Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, and Clayton Foundation Biochemical Institute and Department of Chemistry, The University of Texas, Austin, Texas 78712

Received June 21, 1988; Revised Manuscript Received September 16, 1988

ABSTRACT: Porcine liver DNA polymerase γ contains exonuclease activity capable of digesting DNA in the 3' \rightarrow 5' direction, releasing deoxyribonucleoside 5'-monophosphates. The exonuclease activity excises 3'-terminal bases from both matched and mismatched primer termini, with a preference for mismatched bases. Under polymerization conditions, mismatch excision by the exonuclease occurs prior to polymerization by polymerase γ , and this excision can be inhibited by adding to the reaction a high concentration of dNTP substrates and/or nucleoside 5'-monophosphates. In an M13mp2-based reversion assay for detecting single-base substitution errors, porcine liver polymerase γ is highly accurate; the estimated base substitution error rate is less than one error for each 500 000 bases polymerized. Lower fidelity is observed using reaction conditions that inhibit the exonuclease activity, strongly suggesting that the exonuclease proofreads errors during polymerization. However, in a forward mutation assay capable of detecting all 12 mispairs at a variety of template positions, certain base substitution errors are readily detected even using unperturbed polymerization conditions. Thus, for some errors, polymerase γ is not highly accurate, suggesting that proofreading is not equally active against all mispairs. To examine if the polymerase and exonuclease activities are physically as well as functionally associated, both activities were monitored during purification by four procedures, each based on a different separation principle. The two activities copurify during chromatography using phosphocellulose, heparin-agarose, or double-strand DNA-cellulose, and during velocity sedimentation in a glycerol gradient containing 0.5 M KCl. These results suggest that the polymerase and exonuclease activities are physically associated. It remains to be determined if they reside in the same subunit.

In vertebrate cells, mitochondrial DNA replication is catalyzed by DNA polymerase γ (Pol- γ).¹ This DNA polymerase has been partially purified from a variety of sources, and several of its catalytic properties have been described [for a review, see Fry and Loeb (1987)]. One of the least studied properties of Pol- γ , but one which is critical for a replicative DNA polymerase, is the fidelity of polymerization. The fidelity of DNA synthesis by any DNA polymerase reflects its ability to discriminate against errors as nucleotides are being incorporated onto a primer-template. In addition, certain DNA polymerases contain associated 3' \rightarrow 5' exonuclease activity capable of removing nucleotides from a 3'-OH primer prior to further elongation [for a review, see Kunkel (1988)]. We have recently shown that highly purified preparations of chick embryo Pol- γ contain a proofreading exonuclease activity that substantially enhances the single-base substitution fidelity of this polymerase (Kunkel & Soni, 1988). This observation led to the present study in which we address three questions. (i) Does Pol- γ isolated from a mammalian source contain associated 3' \rightarrow 5' exonuclease activity? (ii) If so, does the exo-

nuclease enhance fidelity by proofreading? (iii) Are the two activities physically associated? The results presented here suggest an affirmative answer to all three questions. In addition, the fidelity measurements required to answer these questions demonstrate that particular base substitution errors occur frequently even under conditions which should permit active proofreading, suggesting that the exonuclease does not efficiently proofread certain mispairs.

EXPERIMENTAL PROCEDURES

Materials

Porcine liver DNA polymerase γ was purified as described (Mosbaugh, 1988; see Results). Rat Novikoff hepatoma DNA polymerase β , fraction VI, was purified as described (Stalker et al., 1976). *Escherichia coli* DNA polymerase I, large (Klenow) fragment, was purchased from Boehringer Mannheim Biochemicals, and avian myeloblastosis virus (AMV) reverse transcriptase was obtained from Pharmacia. The

[†] Parts of this work were supported by Grant GM 32823 from the National Institutes of Health to D.W.M.

^{*} To whom correspondence should be addressed.

[†] National Institute of Environmental Health Sciences.

[§] The University of Texas.

¹ Abbreviations: Pol- β , DNA polymerase β ; Pol- γ , DNA polymerase γ ; Pol- δ , DNA polymerase δ ; Pol I, *E. coli* DNA polymerase I; AMV Pol, avian myeloblastosis virus reverse transcriptase; Hepes, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; monophosphate, nucleoside 5'-monophosphate; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetate.