Articles

Identification of Nucleotide Binding Sites in the Poliovirus RNA Polymerase†

Oliver C. Richards,*,‡ Jeffrey L. Hanson,§ Steve Schultz,§ and Ellie Ehrenfeld‡

Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine, California 92717, and Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309

Received December 20, 1994; Revised Manuscript Received March 13, 1995⊗

ABSTRACT: Poliovirus RNA polymerase ($3D^{pol}$) was cross-linked to [^{32}P]ribonucleoside triphosphates (NTPs) by reduction of oxidized NTP−protein complexes. Cross-linked complexes were digested with cyanogen bromide, and resulting peptides were fractionated by reverse-phase HPLC. ^{32}P-Labeled peptides were purified by secondary HPLC fractionation and/or additional digestion with endoproteinases Glu-C, TPCK−trypsin, or Asp-N followed by another HPLC fractionation. N-Terminal sequences of the major [^{32}P]-peptides were determined, and approximate sizes of these peptides were obtained by SDS−polyacrylamide gel electrophoresis. Two major NTP binding sites in $3D^{pol}$ were found. One site was between Asp-266 and Met-286; possible binding residues in this fragment were Lys-276, Lys-278, or Lys-283. A second binding site was between Ala-57 and Met-74 with Lys-61 or Lys-66 as possible binding residues. Alignment of these regions on the known structure of HIV-1 reverse transcriptase allowed us to predict the position of the downstream nucleotide binding site in the conserved "fingers" subdomain present near the active site cleft of both RNA and DNA polymerases. The N-terminal nucleotide binding site is not contained within a region that is conserved among other polymerases.

The synthesis of RNA from an RNA template is a unique biochemical reaction utilized by RNA viruses to replicate their genomes. The reaction is catalyzed by a class of enzymes, RNA-dependent RNA polymerases, whose biochemical mechanisms and structure−function relationships are not well understood. Recent progress in the structural analyses of other classes of polynucleotide polymerases (DNA-dependent DNA or RNA polymerases; RNA-dependent DNA polymerase) has indicated that all of these enzymes share a limited resemblance to one another (Moras, 1993). Several structural elements have been identified that may account for RNA or DNA substrate specificity (Ollis et al., 1985; Kohlstaedt et al., 1992; Sousa et al., 1993; Delarue et al., 1990), and other conserved residues likely have roles in catalytic activity. Available sequence information indicates that the viral RNA-dependent RNA polymerases will follow in this pattern (Poch et al., 1989).

In the absence of crystallographic data, a variety of methods have been applied to facilitate determination of substrate binding sites by biochemical means. Cross-linking of nucleotides or oligonucleotides to polymerases has been used with some success; examples include photoactivation of derivatized nucleotides (Frischauf & Scheit, 1973; Rush & Konigsberg, 1990; Knoll et al., 1992; Scheng & Dennis, 1993; Hanna et al., 1993), ultraviolet light activation of nucleotides (Pandey et al., 1987; Tirumalai & Modak, 1991; Cheng et al., 1993; Prasad et al., 1993), chemical cross-linking with pyridoxal phosphate (Basu & Modak, 1987; Basu et al., 1988, 1989), and chemical cross-linking with benzaldehyde derivatives of NTPs[1] (Schaffner et al., 1987; Hartmann et al., 1988; Grachev et al., 1989; Riva et al., 1990). These methods are often limited by poor efficiency of cross-linking to polymerases. Another modification, periodate-oxidized NTPs, provides excellent cross-linking efficiency which has been demonstrated for other classes of proteins (Clertant & Cuzin, 1982; Esterbrook-Smith et al., 1976; Gregory & Kaiser, 1979; Lowe & Beechey, 1982; King & Colman, 1983). This method has been employed in the studies described in this report to explore the nucleotide binding site(s) in poliovirus RNA polymerase ($3D^{pol}$).

The poliovirus RNA-dependent RNA polymerase is a 52 kDa protein, encoded in the 3′ end of the poliovirus genome. The protein has been purified from poliovirus-infected cells, and its cDNA has been cloned and expressed in bacterial, insect, and mammalian cells, resulting in the production of active enzyme, functionally identical to the mature protein synthesized during a viral infection (Rothstein et al., 1988; Neufeld et al., 1991). We have previously shown by binding competition studies that the enzyme has a common nucleotide binding site (Richards et al., 1992). Cross-linking of oxidized nucleotides to $3D^{pol}$ showed saturation kinetics, presumably approaching 1 mol of NTP bound/mol of $3D^{pol}$. In preliminary studies utilizing selective proteolysis and analysis of NTP-labeled peptides, a region in $3D^{pol}$ between

† This work was supported by Grant AI 17386 from the National Institutes of Health, and sequencing of peptides was partially supported by Cancer Center Support Grant CA-62203 to the Molecular Biology Shared Resource at the University of California, Irvine.

* To whom correspondence should be addressed.
‡ University of California, Irvine.
§ University of Colorado.

⊗ Abstract published in *Advance ACS Abstracts,* May 1, 1995.

[1] Abbreviations: NTP, nucleoside triphosphate; $3D^{pol}$, poliovirus RNA polymerase; HPLC, high-pressure liquid chromatography; SDS, sodium dodecyl sulfate; cDNA, complementary DNA; TFA, trifluoroacetic acid; PAGE, polyacrylamide gel electrophoresis; PTH, phenylthiohydantoin.

0006-2960/95/0434-6288$09.00/0

Met-189 and Lys-228 was selected as the NTP binding site in the polymerase, assuming a single nucleotide binding site per $3D^{pol}$ molecule.

As part of our efforts to define the catalytic site of $3D^{pol}$, we have extensively explored the NTP binding site, utilizing periodate-oxidized, labeled NTPs and subsequent reduction of the NTP−$3D^{pol}$ complexes. Identification of regions in $3D^{pol}$ to which NTP is bound has been done by proteolysis, peptide purification, and sequencing and size determination of labeled NTP−peptide complexes. This report shows that there are two primary binding sites in $3D^{pol}$, one found in a conserved structural domain common to many polymerases (Poch et al., 1989) and the other near the N-terminus of the protein in a region of unknown structure or function.

EXPERIMENTAL PROCEDURES

Cross-Linking of Oxidized GTP with Poliovirus Polymerase ($3D^{pol}$). Poliovirus $3D^{pol}$ was expressed in *Escherichia coli* harboring pEXC-3D, and crude sonicates were prepared as described by Rothstein et al. (1988). The lysates were clarified by high-speed centrifugation, and polymerase was purified from the supernatant, basically as described elsewhere (Neufeld et al., 1991; Richards et al., 1992). The purification protocol included 0−40% saturation ammonium sulfate precipitation, phosphocellulose (Whatman), Mono Q (Pharmacia LKB), and Phenyl-Superose (Pharmacia LKB) column chromatography, and a final concentration of $3D^{pol}$ by step elution from a second Mono Q column. The product enzyme was greater than 95% pure as shown by electrophoresis in a 10% polyacrylamide−SDS gel and staining of the gel with silver. The specific activity of the enzyme corresponds to the maximum specific activity reported for this protein under similar assay conditions. [α-^{32}P]GTP (Amersham Corp.) was oxidized with sodium periodate, excess periodate was consumed by glycerol, and oxidized GTP was cross-linked to poliovirus $3D^{pol}$ by modification of the procedure of Clertant and Cuzin (1982). The cross-linked mixture contained 2.5 mM oxidized [^{32}P]GTP (0.5 μCi of ^{32}P/μL), 50 mM HEPES, pH 8.0, 1 mM dithiothreitol, 2.5 mM $MgCl_2$, 4−8 nmol of $3D^{pol}$, and 4−10 mM sodium cyanoborohydride in a final volume of 100−150 μL and was incubated at 0 °C in excess of 20 h. Addition of enzyme constituted approximately 50% of the reaction volume and contributed 0.15−0.2 M KCl to the final cross-linking incubation. Excess nucleotides and salts were removed by dialysis against 50 mM NH_4HCO_3, pH 7.8, with frequent medium changes at 4 °C, and the solution was concentrated by vacuum evaporation in the Speed-Vac (Savant).

Preparation and Purification of CNBr-Generated $3D^{pol}$ Peptides. Samples of [^{32}P]GTP−$3D^{pol}$ complexes (60 μL) were adjusted to 5% β-mercaptoethanol, 3 volumes of 15% CNBr (Sigma) in 98% formic acid (Fluka) was added, and digestion ensued in the dark at room temperature for 2−5 h. Samples were diluted to 0.5 mL with water, adjusted to 0.1% trifluoroacetic acid (TFA), and injected directly into a HPLC apparatus. HPLC fractionation utilized a Waters system with a multiwavelength detector. Initial fractionation of CNBr-generated $3D^{pol}$ peptides utilized a Vydac C_4 reverse-phase column (4.6 × 250 mm) which was equilibrated with a mixture of 85% buffer A (0.1% TFA in water) and 15% buffer B (0.1% TFA in acetonitrile). The fractionation was developed with the following regimen: 15% buffer B isocratic wash for 45 min, 15−50% buffer B in 70 min, and 50−100% buffer B in 5 min at a flow rate of 1.0 mL/min. Fractions (1 min) were collected starting immediately after the isocratic wash and were monitored by Cerenkov radiation in the scintillation counter. In some experiments, radioactive peaks were pooled and rerun on a second reverse-phase HPLC column, using a Vydac C_{18} column (2.1 × 250 mm) at a flow rate of 0.3 mL/min. This fractionation was developed with 100% buffer A (0.1% TFA in water) isocratic wash for 35 min, 0−30% buffer B (0.1% TFA in 70% acetonitrile) in 15 min, 30−72% buffer B in 80 min, and 72−100% buffer B in 5 min. Fractions (1 min) were collected after the gradient had reached 30% buffer B and again monitored by Cerenkov radiation.

Secondary Digestion of $3D^{pol}$ Peptides with Endoproteinases. Selected fractions of CNBr-generated peptides were pooled and concentrated to approximately 60 μL by vacuum evaporation in the Speed-Vac at 42 °C and then neutralized with 1 M Na_2CO_3. Thereafter, any one of three endoproteinases was used. Digestion with endoproteinase Glu-C (Boehringer Mannheim) was at room temperature for 4 h followed by another addition of Glu-C to a final estimated ratio of $3D^{pol}$ peptide:Glu-C = 20:1 (mass basis) for further incubation at room temperature overnight. Digestion with TPCK−trypsin (Sigma) was at room temperature overnight to give a final ratio of $3D^{pol}$ peptide:trypsin = 25:1 after two additions of enzyme. Digestion with endoproteinase Asp-N (Boehringer Mannheim) was at room temperature overnight to give a final ratio of $3D^{pol}$ peptide:Asp-N = 30:1 after two additions of Asp-N. After any of the proteolytic digestions, samples were adjusted to 0.5 mL in 20% formic acid and injected directly into the HPLC system. Fractionation was on a Vydac C_{18} reverse-phase column, exactly as described above for the second run of the CNBr-generated peptides.

NH_2-Terminal Sequencing of Peptides. Selected fractions from HPLC were concentrated by vacuum evaporation and sequenced using an Applied Biosystems 470A/900 protein sequencer with an on-line PTH Applied Biosystems 120A analyzer.

Polyacrylamide Gel Analyses of Peptides. Size analysis of small peptides utilized modifications of the procedure of Schagger and von Jagow (1987; Dayhuff et al., 1992). A 10% PAGE−SDS gel [acrylamide:bis(acrylamide) = 29:1] in 1.2 M Tris·HCl, pH 8.45, and 0.1% SDS was overlaid with a stacking gel of 4% acrylamide (29:1), 0.9 M Tris·HCl, pH 8.45, and 0.075% SDS. The running buffer contained 1 M Tris base, 1 M Tricine (Sigma), and 1% SDS, pH 8.1; sample buffer (2×) contained 10% glycerol, 5% β-mercaptoethanol, 3% SDS, 0.125 M Tris·HCl, pH 6.8, and 0.1% bromphenol blue. Gels were run at 30 mA, constant current, until the dye had moved 75% through the resolving gel. Gels were dried directly at 80 °C for 2 h under vacuum to avoid loss of small peptides, and ^{32}P-labeled peptides were detected by autoradiography with a Quanta III screen (Du Pont) at −70 °C.

Undigested $3D^{pol}$ was resolved by standard 10% PAGE−SDS (Laemmli, 1970) and detected both by immunoblotting with anti-$3D^{pol}$ serum and by autoradiography.

Positioning of Cross-Linked Lysine Residues Using the Structure of HIV-1 Reverse Transcriptase. Amino acid sequence alignments are those of Xiong and Eickbush (1990). Sequence alignments for motifs 3−5 of two reverse tran-

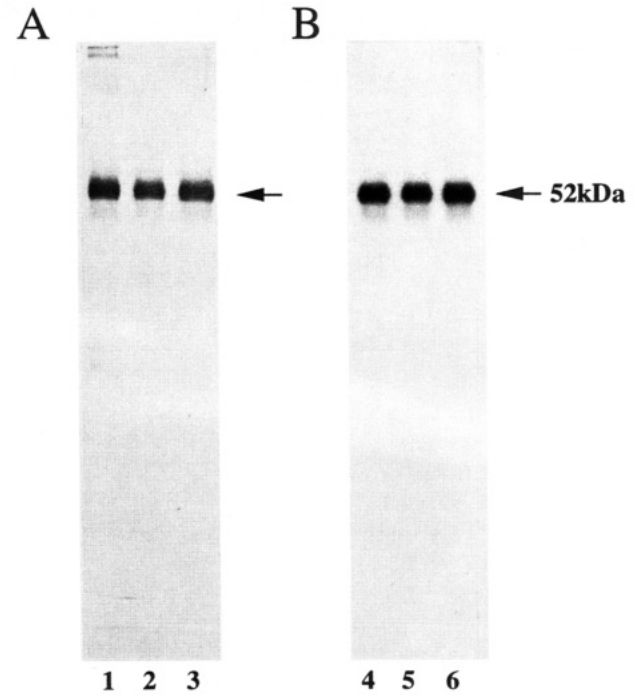

FIGURE 1: Stability of the [^{32}P]GTP−3D^{pol} complex. [^{32}P]GTP was cross-linked to 3D^{pol}, and the complex was equilibrated with 50 mM NH_4HCO_3, pH 7.8. The complex was then incubated at room temperature for 0 h (lanes 1 and 4), 5 h (lanes 2 and 5), or 24 h (lanes 3 and 6), and the incubated samples were electrophoresed in a 10% PAGE−SDS gel. The portion of the gel containing lanes 1−3 was analyzed by Western immunoblotting using anti-3D^{pol} serum. The portion containing lanes 4−6 was dried directly and autoradiographed. The mobility of intact poliovirus 3D^{pol} is indicated by an arrow.

scriptases (HIV-1 and SIV) and of four picornaviral polymerases (poliovirus, encephalomyocarditis virus, foot-and-mouth disease virus, and human rhinovirus-14) were made. The lysine residues implicated in nucleotide binding in poliovirus 3D^{pol} were positioned relative to the invariant glycine in motif 4, assuming no gaps or insertions between these residues. The structure of the polymerase domain of the p66 subunit of HIV-1 reverse transcriptase (Kohlstaedt et al., 1992; Jacobo-Molina et al., 1993) was obtained from the Brookhaven Protein Data Bank.

RESULTS

Characterization of the Cross-Linked NTP−Poliovirus Polymerase Complex and Stability of the Complex. Reduction of periodate-oxidized nucleotide bound to protein produces a stable, cross-linked complex, with no detectable degradation of protein and no loss of bound nucleotide. Figure 1 shows the stability of a [^{32}P]GTP−3D^{pol} complex after dialysis into 50 mM NH_4HCO_3, pH 7.8, to remove unincorporated reagents, incubated at room temperature up to 24 h. Aliquots were removed at 0, 5, and 24 h, and duplicate samples were fractionated in a 10% PAGE−SDS gel. One set of samples was analyzed by immunoblotting (Burnette, 1981; Towbin et al., 1979) with anti-3D^{pol} serum (Figure 1A), and another set was analyzed by direct autoradiography (Figure 1B). Both the ^{32}P content and the amount of immunoreactive material remaining in the 52 kDa band (3D^{pol}) were constant under these conditions, for at least 24 h. Stable cross-linking of NTP to poliovirus RNA polymerase required periodate oxidation of the NTP prior to exposure to the protein and subsequent reduction of the

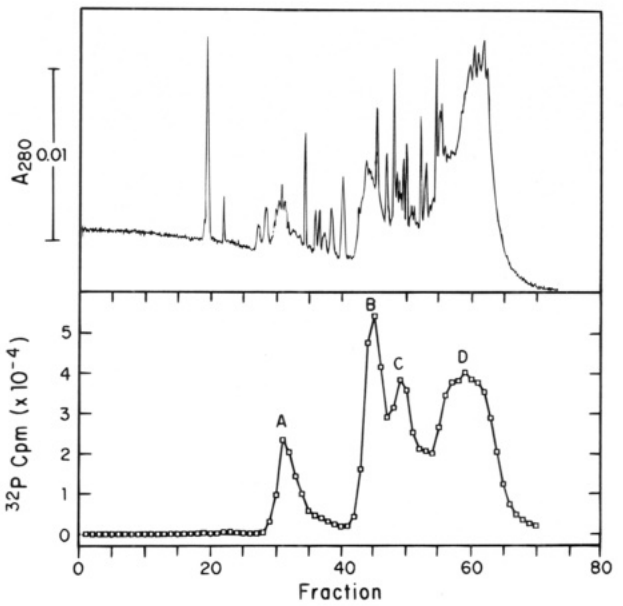

FIGURE 2: Elution profiles obtained during fractionation of cyanogen bromide-digested [^{32}P]GTP−3D^{pol} complexes. A cyanogen bromide digest of 3D^{pol} (4 nmol), cross-linked with [^{32}P]GTP, was fractionated by reverse-phase HPLC; the A_{280} profile is shown in the upper tracing. The radioactivity profile obtained from the HPLC fractions is shown in the lower panel. Peak fractions (A−D) were subsequently analyzed, as described in the text.

oxidized NTP−3D^{pol} complex with $NaCNBH_3$. No detectable complex of 52 kDa was observed in a PAGE−SDS gel by autoradiography of a [^{32}P]NTP-labeled complex in the absence of either treatment (data not shown).

Previous studies (Richards et al., 1992) showed that all four NTPs competed for cross-linking to 3D^{pol}, suggesting that there is a common binding site(s) for all NTPs. Saturation appeared to occur, using the oxidized NTP cross-linking procedure, at 1 mol of GTP/mol of 3D^{pol}. Calculation of the specific activity of cross-linked 3D^{pol} formed in the present study indicated that saturation with NTP was close to 2 mol of GTP/mol of 3D^{pol}. We believe this to be a more accurate figure, based on more accurate estimations of protein levels after cross-linking. Although it was not possible to rule out the presence of some fraction of inactive protein in the enzyme preparation, the previous chromatography of 3D^{pol} on GTP−agarose columns showed generally less than 10% failure to bind (Richards et al., 1992).

Cyanogen Bromide Digestion of the Cross-Linked 3D^{pol} Complex. GTP−3D^{pol} complexes were digested with CNBr as described in Experimental Procedures, and the generated peptides were fractionated by reverse-phase HPLC to give a fairly reproducible A_{280} profile (Figure 2, upper tracing) and a very characteristic ^{32}P profile (Figure 2, lower panel). Material in each peak was subjected to preliminary analysis by polyacrylamide gel electrophoresis and N-terminal sequence analysis. Peaks A and B were selected for further resolution and secondary digestion of labeled peptides, whereas peaks C and D, which contained complex mixtures of multiple peptides, including incomplete digestion products, were not used for further study.

Analysis of Peak A from RP-HPLC. The peak fractions from peak A depicted in the ^{32}P profile from an initial HPLC fractionation were pooled and concentrated by vacuum evaporation and rerun in another reverse-phase HPLC column using a shallower elution gradient; the ^{32}P profile

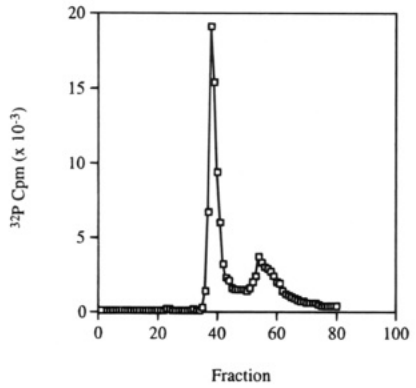

FIGURE 3: Elution profile of radioactivity from the peak A region upon refractionation in HPLC. Cyanogen bromide peptides from the peak A region of a HPLC gradient (see Figure 2) were refractionated in a reverse-phase HPLC C_{18} column, and fractions were assayed for ^{32}P content.

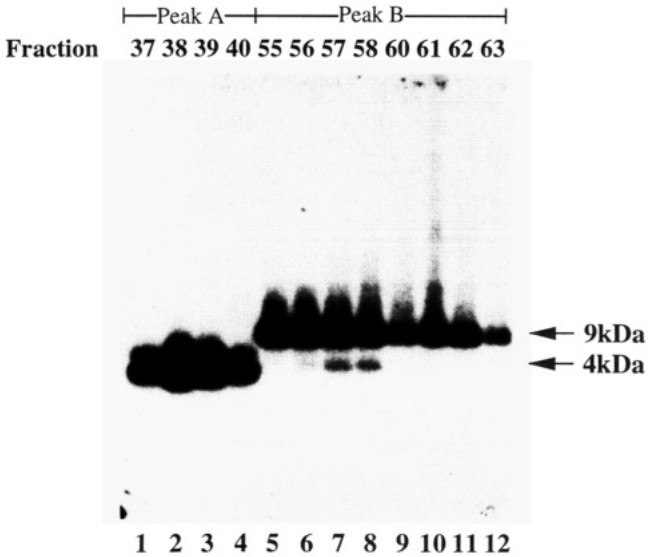

FIGURE 4: Size analysis of cyanogen bromide peptides from HPLC fractions. Aliquots of peak A fractions (see Figure 3) and of peak B fractions (data not shown) from HPLC refractionation of cyanogen bromide-induced peptides were resolved in a 10% PAGE−SDS (Tris−Tricine) gel. The gel was immediately dried and autoradiographed. In the same gel polypeptide size markers (BRL low molecular weight standards) were resolved and detected by silver staining; on the basis of these sizes, the estimated molecular weights of the ^{32}P bands in the autoradiogram are indicated by arrows to the right in the figure. Lanes 1−4 correspond to peak A fractions 37−40 from the HPLC gradient, respectively. Lanes 5−12 correspond to the split peak B fractions 55−58 and 60−63 in a separate HPLC gradient, respectively.

from this run is shown in Figure 3. There was primarily a single ^{32}P peak of [^{32}P]GTP−peptide. Peak fractions (37−40) were separately concentrated, and aliquots were analyzed in a 10% PAGE−SDS (Tris−Tricine) gel to estimate the size of ^{32}P-labeled peptides in these fractions by direct autoradiography of the dried gel. Figure 4, lanes 1−4, shows that the labeled peptide(s) had a mobility estimated at 4 kDa; this would correspond to a peptide of 35−40 residues. N-Terminal sequence analysis of the material in the peak fraction (38) revealed a single peptide that started at Val-252 (after Met-251) (Table 1). The presumed next downstream cleavage site would be at Met-286, generating a peptide of 35 amino acids and consistent with the size predicted by gel analysis. The cross-linking procedure utilized is assumed to cross-linked the GTP to a lysine

Table 1: NH_2-Terminal Sequencing of Peak A [^{32}P]GTP-Labeled Peptide (CNBr Digestion)

Edman cycle	amino acid	yield of PTH (pmol)	residue in 3D^{pol}
1	Val	130	252
2	Leu	136	253
3	Glu	81	254
4	Lys	76	255
5	Ile	121	256

FIGURE 5: Peak A ^{32}P-labeled peptides after various proteolytic digestions. The principal ^{32}P-labeled, cyanogen bromide peptide from peak A is depicted at the top, indicating its observed NH_2 terminus and presumed COOH terminus; lysine positions (K) resident in this peptide are shown. Subsequent endoproteinase Glu-C, trypsin, or Asp-N cleavage of this peptide yielded the principal ^{32}P-labeled peptide shown for each secondary cleavage; remaining lysine positions in these peptides are shown. Trypsin cleavage gave a peptide of undetermined COOH terminus, shown by dashed lines and the broken terminus.

residue, and there are four lysine residues at positions 255, 276, 278, and 283 in this 3D^{pol} peptide (Figure 5). The stoichiometry of GTP binding was estimated from the known specific activity of the GTP and the moles of amino acid residues by sequencing. Approximately 1 mol of GTP/mol of peptide was found for this analysis, confirming that the peptide sequenced was the one which contained [^{32}P]GTP.

Subsequent digestion of peak A material with endoproteinase Glu-C only trimmed this [^{32}P]GTP−peptide by three residues at the NH_2 terminus, as determined by sequencing after isolation by another reverse-phase HPLC (Figure 5 and Table 1). There are no additional Glu residues in this peptide, so we presume the COOH terminus remains Met-286. Trypsin digestion removed Lys-255, but the peptide remained labeled, so Lys-255 does not appear to be the cross-linked residue; however, we do not know the COOH terminus of the tryptic peptide, and we do not know if any of the remaining three Lys residues near the COOH terminus were eliminated. Digestion of peak A with endoproteinase Asp-N and reverse-phase HPLC fractionation yielded peptide with an NH_2 terminus at Asp-266 and a presumed COOH terminus at Met-286. In summary, the analysis of peak A material shows that [^{32}P]GTP binding to 3D^{pol} occurs in the region between Asp-266 and Met-286, most likely at a lysine residue at position 276, 278, or 283.

Sometimes we observed the loss of ^{32}P after secondary proteinase digestions and ensuing HPLC fractionation, as indicated by a decrease in the molar ratio of ^{32}P to peptide recovered. This suggested that some polyphosphate was lost during analysis. Furthermore, when we cross-linked 3D^{pol} with [5,6-^{3}H]UTP, a ratio of 1.2 mol of UTP/mol of 3D was obtained. After CNBr digestion and HPLC fractionation, peak A was obtained. Subsequent trypsin digestion of peak A and HPLC fractionation resulted in a loss of the ^{3}H signal. Thus, there was an apparent loss of the N-base during

Table 2: NH_2-Terminal Sequencing of Peak B [^{32}P]GTP-Labeled Peptide

mode of digestion	Edman cycle	amino acid	yield of PTH (pmol)	residue in 3D
CNBr	1	Arg	19	7
	2	Pro	70	8
	3	Ser	24	9
	4	Lys	41	10
	5	Glu	42	11
	6	Val	49	12
CNBr + Glu-C	1	Ala	70	57
	2	Ile	65	58
	3	Phe	55	59
	4	Ser	20	60
	5	Lys	31	61
CNBr + trypsin	1	Leu	25	50
	2	Lys	20	51
	3	Thr	10	52
	4	Asp	10	53
	5	Phe	13	54
	6	Glu	10	55
	7	Glu	11	56
	8	Ala	11	57
CNBr + Asp-N	1	Asp	52	47
	2	Pro	44	48
	3	Arg	14	49
	4	Leu	43	50
	5	Lys	35	51
	6	Thr	21	52

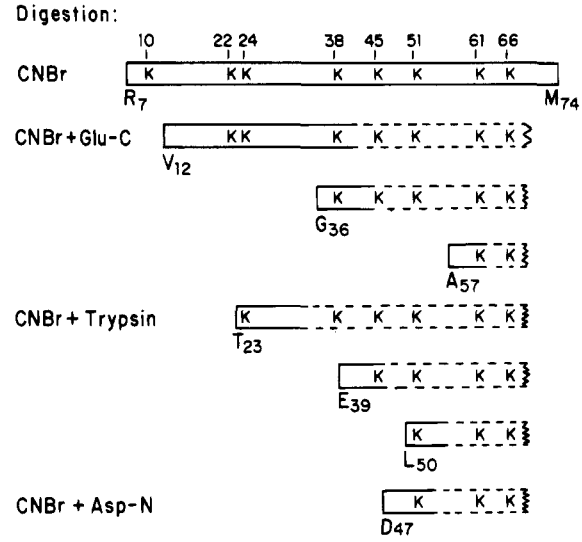

FIGURE 6: Peak B ^{32}P-labeled peptides after various proteolytic digestions. The principal ^{32}P-labeled, cyanogen bromide peptide from peak B is depicted at the top with NH_2 and COOH termini and lysine positions (K) shown. Subsequent endoproteinase cleavages yielded multiple fragments labeled with ^{32}P, and in each case NH_2 termini are indicated. In all secondary cleavages COOH termini were not established, and therefore, remaining lysine residues in these peptides are uncertain; this uncertainty is depicted by dashed lines and a broken terminus.

analysis (data not shown). In a separate experiment nonlabeled, oxidized GTP was cross-linked to 3D^{pol} with [^{3}H]-$NaCNBH_3$ reduction; a complex was obtained with about 2 mol of ^{3}H (1 mol of GTP)/mol of 3D^{pol}. The incorporated ^{3}H appeared to be stable, although precise quantitation was difficult in this experiment. The implication is that the sugar moiety remained stably cross-linked in the complex throughout peptide generation and isolation but that some loss of the base moiety and of the polyphosphate may occur during the manipulations.

Analysis of Peak B from RP-HPLC. Peak B fractions from the initial fractionation of CNBr-induced 3D^{pol} peptides by reverse-phase HPLC were pooled and rerun in a second HPLC fractionation. A double ^{32}P peak was obtained (data not shown). Aliquots of HPLC fractions encompassing the ^{32}P doublet region were analyzed in a 10% PAGE−SDS (Tris−Tricine) gel (Figure 4, lanes 5−12), and there was a major band of about 9 kDa across both peak fractions. N-Terminal sequence analysis of peak fractions from each peak gave a common start at Arg-7 (Table 2). The presumed COOH terminus is Met-74, which would yield a peptide containing 68 amino acid residues. Thus the size of the peptide is consistent with the calculation of the size of the labeled peptide by gel analysis. In this peptide there are eight Lys residues (Figure 6). The stoichiometry for labeled peptide was about 1 mol of GTP/mol of 3D^{pol} peptide. Fractions from the leading edge of the doublet contained an additional labeled peptide of approximately 4 kDa (Figure 4, lanes 5−8). The source of this materials is not known; no second peptide was evident in these fractions by sequence analysis.

Subsequent cleavages of peak B material with endoproteinases TPCK−trypsin, Glu-C, or Asp-N followed by HPLC fractionations and N-terminal sequencing of labeled peptides gave the following results: (1) Digestion with trypsin generated several ^{32}P-labeled peptides, all derived from a common region in the 3D^{pol} structure. These likely represent overlapping, partial digestion products, with the shortest labeled peptide starting at Leu-50 and ending at an unknown site (Table 2). Maximally this peptide would contain three Lys residues (Figure 6). (2) Digestion with Glu-C gave two principal ^{32}P peaks in HPLC and contained overlapping peptides with the shortest, labeled peptide starting at Ala-57 and maximally containing two Lys residues (Table 2; Figure 6). (3) Digestion with Asp-N resulted in overlapping ^{32}P peaks with the shortest, labeled peptide starting at Asp-47. It could contain maximally three Lys residues. The combined data support the conclusion that [^{32}P]GTP binds 3D^{pol} in the vicinity of Lys-61 or Lys-66 as a second NTP binding site in the 3D^{pol} molecule.

Positioning Lysine Residues at 276, 278, and 283 in the Poliovirus 3D Molecule on a HIV-1 Reverse Transcriptase Model. The sequences listed in Figure 7A include motifs 3−5 [or A−C, respectively, from the nomenclature of Poch et al. (1989)] for two reverse transcriptases and four picornaviral RNA polymerases. Of the three potential cross-linking sites between Asp-266 and Met-286 in poliovirus 3D^{pol}, Lys-283 is closest to a well-conserved region (motif 4) whereas Lys-276 and Lys-278 are located near or in a nonconserved region between motifs 3 and 4; indeed, the poliovirus 3D^{pol} contains a 15 amino acid insertion relative to HIV-1 reverse transcriptase immediately upstream of Lys-276. Of the three lysine residues in question, only Lys-278 bears identity to aligned polymerases from two other picornaviruses (encephalomyocarditis virus and foot-and-mouth disease virus).

The positions of the three potential lysine cross-linking sites, as aligned on the structure of HIV-1 reverse transcriptase, are indicated in Figure 7B. The suggested positions of Lys-276, Lys-278, and Lys-283 of poliovirus 3D are placed 13, 11, and 6 residues, respectively, N-terminal to the invariant glycine of motif 4. Notice that these sites are

	Motif 3		Motif 4		Motif 5
SIVmnd	IKKCKRITVL **D** IGDAYFSIPLDPDYRPYTAFTVP	SVNNQ----------AP	GKRYMYNVL PQ**G** WKGSPCIFQGTVASLLEVFRKNH	PTVQ-------	LYQ YM**DD** LFVGSDYTAE
HIV-1	LKKKKSVTVL **D** VGDAYFSVPLDEDFRKYTAFTIP	SINNE----------TP	GIRYQYNVL PQ**G** WKGSPAIFQSSMTKILEPFRKQN	PDIV-------	IYQ YM**DD** LYVGSHLEIG
POLIO(222)	VLMEEKLFAF **D** YT-GYDASLSPAWFEALKMVLEK	IGFGDRVDYIDYLN---HSHHLYKN	KTYCVKGGM PS**G** CS-GTSIFNSMINNLIIRTLLLK	TYKGIDLDHLK	MIA YG**DD** VIASYPHEVD
EMCV	MQGFERVYDV **D** YS-NFDSTHSVAMFRLLAEEFFT	PENGFDPLTREYLESLAISTHAFEE	KRFLITGGL PS**G** CA-ATSMLNTIMNNIIIRAGLYL	TYKNFEFDDVK	VLS YG**DD** LLVATNYYQL
FMDV	FAQYRNVWDV **D** YS-AFDANHCSDAMNIMFEEVFR	TDFGFHPNAEWILKTLVNTEHAYEN	KRITVEGGM PS**G** CS-ATSIINTILNNIYVLYALRR	HYEGVELDTYT	MIS YG**DD** IVVASDYYDL
HRV-14	CLMDGHLMAF **D** YS-NFDASLSPVWFVCLEKVLTK	LGFAGSSLIQSICN----THHIFRD	EIYVVEGGM PS**G** CS-GTSIFNSMINNIIIRTLILD	AYKGIDLDKLK	ILA YG**DD** LIVSYPYELD

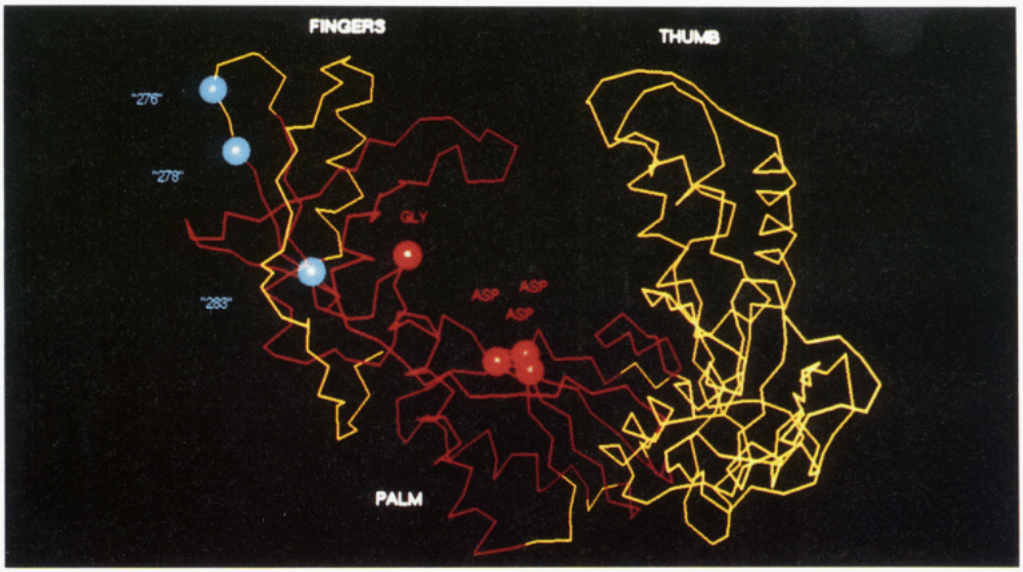

FIGURE 7: Potential poliovirus $3D^{pol}$ cross-linking sites with NTP aligned on the structure of HIV-1 reverse transcriptase. (A, top) The amino acid sequence alignment is that of Xiong and Eickbush (1990). The sequences include motifs 3−5 [or A−C, respectively, from the nomenclature of Poch et al. (1989)]. The sequences of two reverse transcriptases and four picornaviral RNA polymerases are included for comparison. Amino acid residues conserved in RNA-templated polymerases are shown in bold font. The bracketed number to the left of the poliovirus polymerase sequence denotes the amino acid residues N-terminal to this sequence in 3D. The underlined lysine residues in the poliovirus sequence in motif 4 and just upstream of it are the potential cross-linking sites described in this text. (B, bottom) The structure of the polymerase domain of the p66 subunit of HIV-1 reverse transcriptase is represented as an α-carbon trace (Kohlstaedt et al., 1992; Jacobo-Molina et al., 1993). The seven conserved amino acid sequence motifs (Xiong & Eickbush, 1990) are traced in red, and the nonconserved regions are traced in yellow. The red spheres indicate the Asp residues of motifs 3 and 5 that are highly conserved in all classes of polymerases as well as the Gly residue of motif 4 that is invariant in all RNA-dependent polymerases. The positions of the poliovirus $3D^{pol}$ potential cross-linking sites (assuming no gaps or insertions between the Gly residue and the Lys residues) are shown in light blue.

all located in the "fingers" subdomain. Due to the absence of sequence conservation in the region including Lys-276 and Lys-278 and a significant insertion of a peptide segment in picornaviral polymerases N-terminal to Lys-276 relative to HIV-1 reverse transcriptase, a more precise location for these residues cannot be assigned.

The N-terminal portions of polymerase proteins bear no detectable sequence or structure conservation; thus no alignment of the N-terminal nucleotide binding site region in poliovirus $3D^{pol}$, containing Lys-61 and Lys-66, can be performed.

DISCUSSION

Studies of nucleotide binding by a number of DNA and RNA polymerases have contributed significantly to our understanding of the definition of the catalytic pocket, which is thought to represent a conserved motif among all of these enzymes (Ollis et al., 1985; Delarue et al., 1990; Kohlstaedt et al., 1992; Sousa et al., 1993). No comparable information has been obtained about the substrate binding sites for RNA-directed RNA polymerases. Preliminary information has been presented for Qβ RNA polymerase where nucleotide

binding has been shown to be to the virally encoded subunit of the enzyme (Hartmann et al., 1988). In our study we found two primary binding sites in poliovirus $3D^{pol}$. One site is situated between Asp-266 and Met-286 and partially within one of four conserved sequences common to polymerases (Poch et al., 1989). This region contains three Lys residues (out of the total of 38 Lys in the $3D^{pol}$ molecule) at positions 276, 278, and 283. Lysine residues are the presumed acceptor residues for oxidized NTPs in proteins; hence the aforementioned Lys residues in this peptide are likely sites for binding NTP. A second site is found between Ala-57 and Met-74 near the NH_2 terminus and contains two Lys residues at positions 61 and 66. Our previous data (Richards et al., 1992) showed labeled peptides covering both these sites, but we incorrectly assumed a single nucleotide binding site per $3D^{pol}$ molecule. The error likely arose from an overestimate in the protein determination after cross-linking. In some experiments we also observed NTP binding to $3D^{pol}$ downstream of Thr-307 in a peptide derived from peak B after CNBr digestion and HPLC fractionation. Cross-linking to this peptide, however, was inconsistent and may be the result of fortuitous binding to an exposed region of $3D^{pol}$.

Genetic tests of the biochemical predictions from this study have not yet been performed. To date, very few examples with mutations in the NH_2-terminal region of $3D^{pol}$ have been examined. Diamond and Kirkegaard (1994) examined a series of mutations in poliovirus $3D^{pol}$ in the region just upstream and just downstream of the NTP binding locus containing Lys-61 and Lys-66 in this report. For the most part these mutations result in reduced viral RNA accumulations in infected cells, but immediate effects on $3D^{pol}$ activity were not examined. A number of mutations have been constructed in the central portion of $3D^{pol}$, principally in regions of conserved sequences between RNA-templated polymerases. In most cases the mutations are either lethal or yield wild-type polymerase activities. In general, these analyses have not been informative regarding functional effects of structural changes in $3D^{pol}$.

Since all known RNA-dependent polymerases contain several discrete regions of amino acid sequence similarities (suggesting common structural elements) and since the three-dimensional structures of several diverse polymerases including the large (Klenow) fragment of *E. coli* DNA polymerase I, phage T7 RNA polymerase, and HIV-1 reverse transcriptase exhibit strikingly similar structures especially in regions corresponding to the conserved sequence motifs of RNA-dependent polymerases, we decided to utilize this information in attempting to localize the poliovirus 3D NTP cross-linking sites. This analysis included only the downstream cross-linked segment that contains Lys-276, Lys-278, and Lys-283 since this region is located near conserved sequence motifs. The N-terminal cross-linked fragment exhibits no amino acid sequence homologies with HIV-1 reverse transcriptase. On the basis of the previously described sequence alignments, Lys-276, Lys-278, and Lys-283 were positioned in the structure of the HIV-1 reverse transcriptase since this is the most closely related polymerase for which a high-resolution three-dimensional structure is available. As is apparent in Figure 7B, all three lysine residues are located in the fingers subdomain. Lys-283 is perhaps the most reliably positioned since it is closest to the highly conserved residues of motif 4. The Lys-283 side chain could easily reach into the active site cleft at the base of the fingers domain. In contrast, Lys-276 and Lys-278 are on the back side of the fingers subdomain, well away from the active site cleft (Figure 7B). However, we should not rule out these residues as the one(s) that cross-link(s) to the NTP since they are in a region that is poorly conserved with regard to the other polymerases, and therefore, the structure of the poliovirus polymerase may be quite different from that of HIV-1 reverse transcriptase in this region of the structure. On the basis of their location relative to motifs 3 and 4, Lys-276 and Lys-278 in the poliovirus polymerase structure will most likely be in the tip region in the fingers subdomain. Whereas at first glance the tip of the fingers may appear to be well removed from the proposed active site region of the polymerase cleft (which is thought to be near the base of the fingers subdomain), many of the mutations in HIV-1 that give rise to azidothymidine resistance are located in the tip region of the fingers subdomain, suggesting that this region may be important for NTP binding.

A chemical instability of the bound nucleotide to $3D^{pol}$ was observed in this study. We observed a decrease in the ratio of [^{32}P]NTP per molecule of $3D^{pol}$ with increased manipulation of peptide during analyses, particularly at some stage(s) of neutralization of HPLC fractions after initial fractionation of CNBr-generated peptides, during proteolytic digestions, or during secondary HPLC fractionations. Loss of ^{32}P relative to that in the initial NTP−3D complex is probably due to polyphosphate loss by β-elimination. Previous evidence has been presented for lability of the polyphosphate region of NTP in a pH-dependent and temperature-dependent reaction. Rayford et al. (1985) have observed selective loss of polyphosphate in a model system, and Lowe and Beechey (1982) and King and Colman (1983) have examined polyphosphate loss in some detail using oxidized NTPs in protein cross-linking studies. We attempted to minimize, but did not exclude, this reaction by performing the cross-linking at 0 °C and at pH 8.0, to facilitate reaction with uncharged ϵ-NH_2 residues of lysine in $3D^{pol}$. We also observed a loss of ^3H from [^{3}H]UTP-cross-linked material during generation and purification of peptides, likely due to cleavage at the N-glycosidic linkage. [^{3}H]UTP labeling was performed only once; thus the data are insufficient to discern any pattern, and no mechanism(s) for the apparent instability of cross-linked nucleotide has (have) been suggested.

ACKNOWLEDGMENT

We are grateful to Bob Schackmann, University of Utah Core Facility, for initial sequencing and HPLC assistance on this project.

REFERENCES

Basu, A., & Modak, M. J. (1987) *Biochemistry 26*, 1704−1709.
Basu, A., Nanduri, V. B. Gerard, G. F., & Modak, M. J. (1988) *J. Biol. Chem. 263*, 1648−1653.
Basu, A., Tirumalai, R. S., & Modak, M. J. (1989) *J. Biol. Chem. 264*, 8746−8752.
Burnette, W. N. (1981) *Anal. Biochem. 112*, 195−203.
Cheng, N., Merrill, B. M., Painter, G. R., Frick, L. W., & Furman, P. A. (1993) *Biochemistry 32*, 7630−7634.
Clertant, P., & Cuzin, F. (1982) *J. Biol. Chem. 257*, 6300−6305.
Dayhuff, T. J., Gesteland, R. F., & Atkins, J. F. (1992) *BioTechniques 13*, 500−503.

Delarue, M., Poch, O., Tordo, N., Moras, D., & Argos, P. (1990) *J. Protein Eng. 10*, 461−467.
Diamond, S. E., & Kirkegaard, K. (1994) *J. Virol. 68*, 863−876.
Esterbrook-Smith, S. B., Wallace, J. C., & Keech, D. B. (1976) *Eur. J. Biochem. 62*, 125−130.
Frischauf, A. M., & Scheit, K. H. (1973) *Biochem. Biophys. Res. Commun. 53*, 1227−1233.
Grachev, M. A., Lukhtanov, E. A., Mustaev, A. A., Zaychikov, E. F., Abdukayumov, M. N., Rabinov, I. V., Richter, V. I., Skoblov, Y. S., & Chistyakov, P. G. (1989) *Eur. J. Biochem. 180*, 577−585.
Gregory, M. R., & Kaiser, E. T. (1979) *Arch. Biochem. Biophys. 196*, 199−208.
Hanna, M. M., Zhang, Y., Reidling, J. C., Thomas, M. J., & Jou, J. (1993) *Nucleic Acids Res. 21*, 2073−2079.
Hartmann, G. R., Biebricher, C., Glaser, S. J., Grosse, F., Katzameyer, M. J., Lindner, A. J., Mosig, H., Nasheuer, H.-P., Rothman-Denes, L. B., Schaffner, A. R., Schneider, G. J., Stetter, K.-O., & Thomm, M. (1988) *Biol. Chem. Hoppe-Seyler 369*, 775−788.
Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Lu, X., Tantillo, C., Williams, R. L., Kamer, G. Ferris, A. L., Clark, P., Hizi, A., Hughes, S. H., & Arnold, E. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 6320−6324.
King, M. M., & Colman, R. F. (1983) *Biochemistry 22*, 1656−1665.
Knoll, D. A., Woody, R. W., & Woody, A.-Y. M. (1992) *Biochim. Biophys. Acta 1121*, 252−260.
Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., & Steitz, T. A. (1992) *Science 256*, 1783−1790.
Laemmli, U. K. (1970) *Nature 227*, 680−685.
Lowe, P. N., & Beechey, R. B. (1982) *Biochemistry 21*, 4073−4082.
Moras, D. (1993) *Nature 364*, 572−573.
Neufeld, K. L., Richards, O. C., & Ehrenfeld, E. (1991) *J. Biol. Chem. 266*, 24212−24219.
Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., & Steitz, T. A. (1985) *Nature 313*, 762−766.
Pandey, V. N., Williams, K. R., Stone, K. L., & Modak, M. J. (1987) *Biochemistry 26*, 7744−7748.
Poch, O. Sauvaget, I., Delarue, M., & Tordo, N. (1989) *EMBO J. 8*, 3867−3874.
Prasad, R., Kumar, A., Widen, S. G., Casas-Finet, J. R., & Wilson, S. H. (1993) *J. Biol. Chem. 268*, 22746−22755.
Rayford, R., Anthony, D. D., O'Neill, R. E., & Merrick, W. C. (1985) *J. Biol. Chem. 260*, 15708−15713.
Richards, O. C., Yu, P., Neufeld, K. L., & Ehrenfeld, E. (1992) *J. Biol. Chem. 267*, 17141−17146.
Riva, M., Carles, C., Sentenac, A., Grachev, M. A., Mustaev, A. A., & Zaychikov, E. F. (1990) *J. Biol. Chem. 265*, 16498−16503.
Rothstein, M. A., Richards, O. C., Amin, C., & Ehrenfeld, E. (1988) *Virology 164*, 301−308.
Rush, J., & Konigsberg, W. H. (1990) *J. Biol. Chem. 265*, 4821−4827.
Schaffner, A. R., Jorgensen, E. D., McAllister, W. T., & Hartmann, G. R. (1987) *Nucleic Acids Res. 15*, 8773−8781.
Schagger, H., & von Jagow, G. (1987) *Anal. Biochem. 166*, 368−379.
Sheng, N., & Dennis, D. (1993) *Biochemistry 32*, 4938−4942.
Sousa, R., Chung, Y. J., Rose, J. P., & Wang, B.-C. (1993) *Nature 364*, 593−599.
Tirumalai, R. S., & Modak, M. J. (1991) *Biochemistry 30*, 6436−6443.
Towbin, H., Staehelin, T., & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4350−4354.
Xiong, Y., & Eickbush, T. H. (1990) *EMBO J. 9*, 3353−3362.

BI942921Q