Design and Synthesis of RNA Miniduplexes via a Synthetic Linker Approach†

Michael Y.-X. Ma,* Lorne S. Reid,‡ Shane C. Climie, Wing C. Lin, Raya Kuperman, Martin Sumner-Smith, and Richard W. Barnett

Allelix Biopharmaceuticals Inc., 6850 Goreway Drive, Mississauga, Ontario L4V 1P1, Canada

Received October 2, 1992; Revised Manuscript Received November 17, 1992

ABSTRACT: Double-stranded oligodeoxyribonucleotides or single-stranded oligoribonucleotides with specific secondary structure have been proposed as potential antagonists to target nucleic acid-binding proteins (the sense approach). A major limitation of this strategy is that these derivatives are generally considered to be too large for pharmaceutical applications. We have developed a synthetic linker approach whereby nucleic acid duplexes of a much smaller size (miniduplexes) can be generated directly from a standard oligonucleotide synthesis. In this approach, four synthetic linkers (derivatized respectively from 1,9-nonanediol, triethylene glycol, 1,3-propanediol, and hexaethylene glycol) of different length and hydrophobicity were designed and incorporated into a model RNA molecule based on the TAR stem–loop structure of HIV-1. Their thermal stabilities were evaluated by measuring denaturation profiles (T_m measurements). These linker-derivatized RNA molecules were then assessed for their ability to bind to either a full-length protein (HIV-1 Tat protein) or a short peptide (Tat-derived peptide) through RNA mobility shift assays. Results from this study indicate that such modified miniduplex structures retain full binding activity relative to that of the wild-type sequence (K_d values), while T_m values were increased by 24–31 °C compared to an open duplex of the same length. This system provides a new direction in the use of nucleic acid miniduplexes as a novel class of oligonucleotide analogues for both fundamental research and possible therapeutic applications.

Synthetic oligonucleotides are being considered as potential therapeutic agents for artificial regulation of gene expression. With the "antisense approach" and the "antigene approach", specific nucleic acid sequences are targeted against either single-stranded messenger RNA or double-stranded genomic DNA [for recent reviews, see Uhlmann and Peyman (1990) and Helene and Toulme (1990)]. However, with the "sense approach" nucleic acid–protein interactions are targeted. The sense approach is a potentially viable alternative since virtually all biological functions of nucleic acids (e.g., transcription, translation, splicing, or degradation) involve a sequence-specific interaction with a protein. A double-stranded DNA or single-stranded RNA molecule with a specific secondary structure could be used as an antagonist to specifically compete with the natural nucleic acid substrate of a binding protein. This hypothesis has recently been tested using double-stranded phosphorothioate oligodeoxynucleotides (21–39 bp) containing either the octamer or κB consensus sequences, respectively, to interfere with a transcription factor or a nuclear factor (Bielinska et al., 1990). In vivo, double-stranded DNA is much more stable than its single-stranded homologue (Harel-Bollan et al., 1989). The challenge resides in reducing such nucleic acid duplexes to a pharmaceutically acceptable size (e.g., 6–8 bp) while still retaining their biological activity.

When two single-stranded nucleic acid molecules (6–8 bp) anneal, the resulting duplex is usually not stable under physiological conditions. We refer to this class of nucleic acid structures as "miniduplexes". The melting temperatures of such miniduplexes are estimated to be between 20 and 30 °C under physiological conditions (Itakura et al., 1984). This thermal instability and the difficulty in forming a stable complex between two individual molecules preclude their practical use. In nature, short duplex segments with a sequence-specific biological activity (e.g., protein binding) are generally buried within a much longer duplex or two single strands are linked via a nucleotide loop. Introduction of a nucleotide loop (four nucleotides or greater), under certain circumstances, may significantly alter the proximal duplex structure (Antao et al., 1991) and may increase the synthetic cost and complexity of the miniduplex. Several approaches to the synthesis of stable nucleic acid miniduplexes have been described. Common strategies involve the incorporation of a covalently-closed base pair at one end of the duplex by means of a presynthesized, coupled base pair (Petric et al., 1991), or via a postsynthetic cross-linking step (Matteucci & Webb, 1987; Cowart & Benkovic, 1991; Ferentz & Verdine, 1991). These approaches are labor-intensive and result in low yields. In addition, such covalently "locked in" duplexes lack the flexibility which may be required for a conformational change upon protein binding (Petric et al. 1991). More recently, the synthesis and physicochemical properties of a palindromic DNA duplex linked through a hexaethylene glycol chain has been reported (Durand et al., 1990). However, the nature and the optimum characteristics of the bridging chain were not determined and the modified oligonucleotides were not shown to have any biological function.

We have developed a novel synthetic linker approach for generating a sequence-specific duplex structure with a pre-selected biological function. A series of synthetic linkers of different length and hydrophobicity have been integrated into a model RNA molecule in an automated fashion with high efficiency. These linker-derivatized RNA analogues were assessed by gel mobility shift assays for their ability to bind to either a full-length protein or a short peptide. Results from this study clearly indicate that the miniduplexes have enhanced thermal stability and retained biological activity.

† Parts of this work were presented during "The 1991 San Diego Conference on Nucleic Acids: The Leading Edge", November 20–22, 1991, San Diego, CA, and at "Recognition Studies in Nucleic Acids II", April 12–15, 1992, Sheffield, England.

* To whom correspondence should be addressed.

‡ This paper is dedicated to the fond memory of our dear friend and colleague, Dr. Lorne S. Reid, who died suddenly December 17, 1992.

0006-2960/93/0432-1751$04.00/0

MATERIALS AND METHODS

Computer Modeling of Optimum Length for Synthetic Linkers. All structures were modeled with QUANTA/CHARMm software (Rev 3.2.1) from Polygen Corp. on a Silicon Graphics 4D25TG workstation. A 4-mer duplex of DNA was modeled in canonical B form. Partial charges were based upon the DNAH RTF files of CHARMm version 21.0. Alkane chains of varying length or composition were built between the 5′-O atom of the A chain and the 3′-O atom of the B chain. Partial charges on the alkane chain were assigned with the rule-based method of the model builder. The atoms of the nucleic acid duplex were constrained to remain fixed in space, and CHARMm minimization was then applied to move the linker atoms into a low energy configuration (the cutoff for nonbonded interactions was switched on at 14 Å, and off at 17.5 Å, with a distance-dependent dielectric). The Associated Basis Set Newton Raphson (ABNR) algorithm was utilized to minimize the structures to convergence (Δrmsf <0.0001 kcal/(mol Å)).

Synthesis of Linker-Derivatized Phosphoramidites **3a–d**. The modified phosphoramidites **3a–d** were prepared via a two-step procedure: dimethoxytritylation and phosphitylation. All chemical reagents were purchased from Aldrich Chemical Co. Analytical TLC[1] was performed on silica gel 60 F_{254} plates (Merck) and developed in one of the following eluents: system A [$CHCl_3$/MeOH (20:1 v/v)], or system B [petroleum ether/EtOAc/TEA (50:10:1 v/v/v)]. Nuclear magnetic resonance (NMR) spectra were recorded on a Bruker AM-500 or Bruker AM-300 spectrometer with TMS as internal reference for ^1H spectra and 85% phosphoric acid as external reference for ^{31}P spectra. Chemical shifts are given in parts per million (ppm) and signals are described as s (singlet), d (doublet), t (triplet), or m (multiplet).

General Procedure for Dimethoxytritylation of Compounds **1a–d**. The starting diol compound (30 mmol) **1a–d** was coevaporated twice with anhydrous pyridine. The residue was then dissolved in fresh, dry pyridine (110 mL) to yield a final diol concentration of about 1 mmol/5 mL. Next, 5 g of 4,4′-dimethoxytrityl chloride (DMT-Cl) (20 mmol) was added in small portions under argon. The reaction was followed at room temperature by TLC (MeOH/$CHCl_3$, 1:9 v/v) until the appearance of a product spot that was intense relative to the remaining DMT-Cl. The reaction was generally complete after 2–4 h. Upon completion, 25 mL of MeOH was added to quench excess DMT-Cl, and the mixture was stirred for an additional 15 min. The solution was then concentrated to a syrup under vacuum, and the residue was resuspended in 100 mL of $CHCl_3$. The chloroform phase was washed once with 5% $NaHCO_3$ (50 mL) and twice with saturated NaCl solution (brine). The aqueous phase was back-extracted with $CHCl_3$ (50 mL). The organic phases were combined and dried over anhydrous sodium sulfate. After filtration, the solution was evaporated to an oily residue under reduced pressure. This residue was purified by flash chromatography on silica gel. The column was first eluted with petroleum ether/EtOAc (5:1 v/v) and then with petroleum ether/EtOAc (2:1 v/v). Fractions containing the final product were combined and the solvent removed. The resulting oil was dried overnight under high vacuum. The isolated yields, based on the starting amount of DMT-Cl, ranged from 60 to 80%. Products were characterized by TLC and ^1H NMR.

1-O-(4,4′-Dimethoxytrityl)-1,9-nonanediol (**2a**). TLC (silica gel, system A): R_f 0.76. ^1H NMR ($CDCl_3$, 500 MHz): δ 1.25–1.60 (15 H, m, CH_2, OH); 3.03 (t, 2 H, DMTOCH_2); 3.58–3.63 (m, 2 H, CH_2OH); 3.77 (s, 6 H, OCH_3); 6.82 (d, J = 8 Hz, 4 H, arom H ortho of OCH_3); 7.15–7.45 (m, 9 H, arom. H).

1-O-(4,4′-Dimethoxytrityl)-3,6-dioxa-1,8-octanediol (**2b**). TLC (silica gel, system A): R_f 0.64. ^1H NMR ($CDCl_3$, 500 MHz): δ 2.38 (s, 1 H, OH); 3.22 (t, 2 H, J = 5 Hz, DMTOCH_2); 3.60–3.90 (m, 16 H, OCH_2CH_2O); 6.78–6.85 (m, 4 H, arom H ortho of OCH_3); 7.18–7.50 (m, 9 H, arom. H).

1-O-(4,4′-Dimethoxytrityl)-1,3-propanediol (**2c**). TLC (silica gel, system A): R_f 0.79. ^1H NMR ($CDCl_3$, 500 MHz): δ 1.80–1.88 (m, 2 H, $CH_2CH_2CH_2$); 2.16 (s, 1 H, OH); 3.27 (t, 2 H, J = 6 Hz, DMTOCH_2); 3.71–3.80 (m, 2 H, CH_2CH_2OH); 3.80 (6 H, s, OCH_3); 6.80–6.85 (m, 4 H, arom H ortho of OCH_3); 7.18–7.45 (m, 9 H, arom. H).

1-O-(4,4′-Dimethoxytrityl)-3,6,9,12,15-pentaoxaheptadecanediol (**2d**). TLC (silica gel, system A): R_f 0.39. ^1H-NMR ($CDCl_3$, 500 MHz): δ 3.01 (s, 1 H, OH); 3.23 (t, 2 H, J = 5 Hz, DMTOCH_2); 3.54–3.80 (m, 28 H, OCH_2CH_2O); 6.78–6.83 (m, 4 H, arom H ortho of OCH_3); 7.15–7.47 (m, 9 H, arom H).

General Procedure for Phosphitylation of Compounds **2a–d**. The respective monotritylated compound **2a–d** (5 mmol) was dissolved in dry THF (50 mL). Anhydrous diisopropylethylamine (DIPEA) (20 mmol) was injected under a weak flow of argon. The phosphitylating reagent, 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (10 mmol), was added via a syringe over a period of 2–5 min. A white precipitate was quickly formed. The reaction mixture was stirred at room temperature for 1–2 h and monitored by TLC (system B). When the reaction had gone to completion, the excess of phosphitylating reagent was quenched by adding several ice cubes. The mixture was diluted with ethyl acetate (150 mL) and triethylamine (5 mL). The solution was transferred to a separatory funnel and extracted once with 10% aqueous sodium carbonate and twice with brine. The organic phase was dried over anhydrous sodium sulfate, filtered, and evaporated to dryness under reduced pressure. The residue was purified by flash chromatography on silica gel using petroleum ether/EtOAc/TEA (20:10:1 v/v) as eluent. Fractions containing pure product were combined, evaporated, and dried overnight under high vacuum to remove traces of TEA. The product was stored at −20 °C. The yields of isolated products varied from 65 to 80%. The products were characterized by TLC, ^1H NMR, and ^{31}P NMR.

1-O-(4,4′-Dimethoxytrityl)-9-O-[(N,N-diisopropylamino)-β-cyanoethoxyphosphino]-1,9-nonanediol (**3a**). TLC (silica gel, solvent B): R_f 0.84. ^1H NMR ($CDCl_3$, 500 MHz): δ 1.14–1.62 [26 H, m, CH_2, $CH(CH_3)_2$]; 2.63 (t, 2 H, J = 6.5 Hz, CH_2CN); 3.02 (t, 2 H, J = 6.5 Hz, DMTOCH_2); 3.54–3.88 [2 m with one s centered at 3.78, 12 H, OCH_3, CH_2OP, POCH_2CH_2CN, NCH(CH_3)$_2$]; 6.79–6.84 (m, 4 H, arom H ortho of OCH_3); 7.17–7.45 (m, 9 H, arom. H). ^{31}P NMR ($CDCl_3$, 121 MHz): 122.4 ppm.

1-O-(4,4′-Dimethoxytrityl)-8-O-[(N,N-diisopropylamino)-β-cyanoethoxyphosphino]-3,6-dioxa-1,8-octanediol (**3b**). TLC (silica gel, solvent B): R_f 0.48. ^1H NMR ($CDCl_3$, 500 MHz): δ 1.13–1.18 [12 H, 2d, $CH(CH_3)_2$]; 2.51–2.64 (m, 2 H, CH_2CN); 3.23 (t, 2 H, J = 5 Hz, DMTOCH_2); 3.56–3.86

[1] Abbreviations: BSA, bovine serum albumin; DIPEA, *N,N*-diisopropylethylamine; DMAP, 4-dimethylaminopyridine; bp, base pair; DMT, 4,4′-dimethoxytrityl; EtOAc, ethyl acetate; HIV-1, human immunodeficiency virus type 1; nt, nucleotide; TBAF, tetrabutylammonium fluoride; TLC, thin-layer chromatography; T_m, thermal denaturation temperature; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; TEA, triethylamine.

[m, 20 H, OCH_3, OCH_2CH_2O, CH_2OP, $POCH_2CH_2CN$, $NCH(CH_3)_2$]; 6.77–6.86 (m, 4 H, arom. H ortho of OCH_3); 7.18–7.47 (m, 9 H, arom H). ^{31}P NMR ($CDCl_3$, 121 MHz): 148.6 ppm.

1-O-(4,4'-Dimethoxytrityl)-3-O-[(N,N-diisopropylamino)-β-cyanoethoxyphosphino]-1,3-propanediol (**3c**). TLC (silica gel, solvent B): R_f 0.79. ^1H NMR ($CDCl_3$, 500 MHz): δ 1.00–1.30 [12 H, 2d, $CH(CH_3)_2$]; 1.89–1.97 (m, 2 H, $CH_2CH_2CH_2$); 2.44–2.51 (m, 2 H, CH_2CN); 3.14–3.19 (m, 2 H, $DMTOCH_2$); 3.50–3.88 [m, 12 H, OCH_3, CH_2OP, $POCH_2CH_2CN$, $NCH(CH_3)_2$]; 6.74–6.84 (m, 4 H, arom. H ortho of OCH_3); 7.14–7.47 (m, 9 H, arom H). ^{31}P NMR ($CDCl_3$, 121 MHz): 147.3 ppm.

1-O-(4,4'-Dimethoxytrityl)-17-O-[(N,N-diisopropylamino)-β-cyanoethoxyphosphino]-3,6,9,12,15-pentaoxaheptadecanediol (**3d**). TLC (silica gel, solvent B): R_f 0.12. ^1H NMR ($CDCl_3$, 500 MHz): δ 1.15–1.21 [12 H, 2d, $CH(CH_3)_2$]; 2.57–2.66 (m, 2 H, CH_2CN); 3.23 (t, 2 H, J = 5 Hz, $DMTOCH_2$); 3.56–3.91 [m, 32 H, OCH_3, OCH_2CH_2O, CH_2OP, $POCH_2CH_2CN$, $NCH(CH_3)_2$]; 6.76–6.85 (m, 4 H, arom H ortho of OCH_3); 7.16–7.48 (m, 9 H, arom H). ^{31}P NMR ($CDCl_3$, 121 MHz): 148.6 ppm.

Oligoribonucleotide Synthesis. Controlled pore glass (CPG) was used as the solid support matrix. Oligoribonucleotides (RNA) were prepared according to the method of Ogilvie et al. (1988), employing 5'-dimethoxytrityl-2'-*tert*-butyldimethylsilyl-ribonucleoside 3'-CE-phosphoramidites (ChemGenes Corp., MA, or Peninsula Labs, CA). Syntheses were carried out on an Applied Biosystems 380B synthesizer using a modified 0.2-μmol cycle (Andrus, 1988). To synthesize linker-containing oligoribonucleotides, the appropriate linker-derivatized phosphoramidite **3a–d** (dissolved in dry acetonitrile, 0.2–0.3 M) was coupled to the support-bound oligonucleotide at the desired location, using the same synthetic cycle as for standard nucleoside phosphoramidites. The average coupling yield, as assayed by trityl measurement, was 94–96%.

Cleavage from the support, base and phosphate deprotection, and removal of the 2'-O-TBDMS group were performed by established procedures (Scaringe et al., 1990). The crude oligonucleotides in TBAF solution were desalted on a C_{18} Sep-Pak cartridge prior to standard electrophoretic purification using 15–20% polyacrylamide/7 M urea gels. Product bands were visualized by UV-shadowing, cut out, and eluted from the gel matrix. The eluted oligomers were finally desalted on a C_{18} Sep-Pak cartridge and quantified by OD_{260} measurements. Each oligonucleotide linker conjugate was checked for homogeneity and "sized" by 5'-^{32}P-end labeling/analytical PAGE against the crude material and oligonucleotide markers. These RNA oligomers were further characterized by enzymatic RNA sequencing (Donis-Keller, 1980) or base-composition analysis (Seela & Kaiser, 1987).

Gel Electrophoresis and RNA Mobility Shift Assay. Linker-derivatized oligoribonucleotides (**5A–5CC**) and the control sequences (**4**, **6**, and **7**) were 5'-^{32}P-labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP. The labeled oligomers were then purified by phenol/chloroform extraction/EtOH precipitation or spin-column filtration (Bio-Rad, Bio-Spin 30). Prior to binding assays, the RNAs were dissolved in 20 mM Tris-HCl (pH 7.5)/100 mM NaCl, heated to 85 °C for 3 min, then slow-cooled to room temperature. Binding assays were carried out in 20-μL reaction mixtures containing 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM DTT, 1 mM EDTA, 0.5 unit/mL RNasin (Promega), 0.09 μg/mL BSA, 5% (v/v) glycerol, 0.1 nM ^{32}P-labeled RNA (2000–5000 cpm), and either peptide derived from the HIV-1 Tat protein

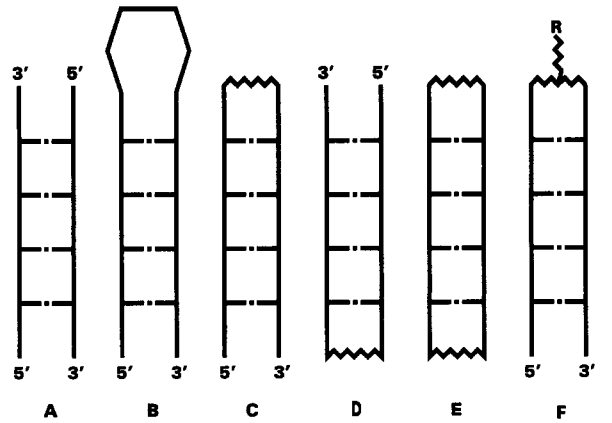


FIGURE 1: Schematic representation of the synthetic linker approach. A relatively unstable nucleic acid miniduplex (structure A) is normally stabilized, in nature, by a naturally-occurring nucleotide loop (structure B). This duplex stabilizing effect can be similarly achieved via a synthetic linker (denoted as zig-zag lines) by covalently connecting the ends of two complementary single strands either from the top of the duplex (structure C), the bottom of the duplex (structure D), or both (structure E). Reactive functionalities (denoted as the group R) useful for oligonucleotide derivatization can also be incorporated into the synthetic linker through a secondary tether bridge (structure F).

RKKRRQRRRPPQGS (amino acids 49–62 of HIV LAI isolate) (Weeks et al., 1990; Delling et al., 1991) (American Peptide Co., Santa Clara, CA) or full-length Tat protein (American Bio-Technologies, Inc.) at a concentration of 0.5 pM to 1000 nM as indicated in the figure legends (Roy et al., 1990a). The reactions were incubated at 23 °C for 25 min, chilled on ice for 5 min, then loaded on 5% native polyacrylamide gels (acrylamide:bisacrylamide = 30:0.8 w/w) containing 5% glycerol. The gels were prerun for 15 min prior to loading and then run for 2.5 h at a constant current of 30 mA at 4 °C in 0.5× TBE buffer. The gels were dried onto DEAE paper (Whatman DE81) and exposed to Kodak X-Omat X-ray film with an intensifying screen overnight at −70 °C. Competition binding experiments were carried out as described above except that the concentration of Tat protein was kept constant at 100 nM and unlabeled competitor RNA was added in a concentration range of 0.9 nM to 5000 nM.

Thermal Denaturation Experiments. T_m measurements were carried out in 100 mM NaCl/10 mM sodium phosphate buffer (pH 7.0). Samples were heated from 25 to 85 °C in 1 °C increments using a HP 8459 UV/VIS spectrophotometer and a HP 89100A temperature controller. The concentration of nucleic acid was 2.5–3.0 μM, and absorbance was monitored at 260 nm. T_m values were determined by a first-derivative plot of absorbance vs temperature. Each experiment was performed in duplicate and the average was reported as the thermal denaturation temperature.

RESULTS

Design and Synthesis of Synthetic Linkers. A naturally-occurring nucleotide loop (Figure 1, structure B) may add substantial stability to an unlinked open duplex (Figure 1, structure A). However, the additional synthetic cost, plus possible interference of the loop with the duplex itself or with a nucleic acid-binding protein, typically renders this choice impractical for the development of a pharmaceutical agent. A synthetic linker, covalently connecting the 3'- and 5'-ends of two complementary single strands, will prevent the dissociation of the resulting duplex. Moreover, by incorporating synthetic linkers at both ends of a given duplex, a covalently-closed cyclic duplex is formed (Figure 1, structure

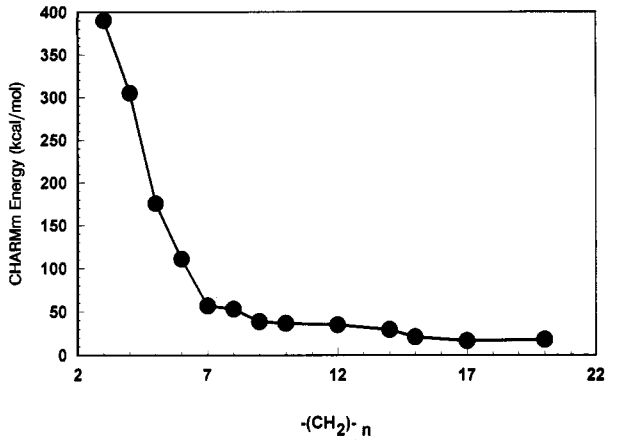


FIGURE 2: Computer modeling of the optimum chain length for synthetic linkers. CHARMm Energy indicates the total calculated energy, and *n* indicates the number of methylene groups between the terminal phosphates. In order to minimize electrostatic repulsion between the terminal phosphates while still enabling the duplex to form, *n* should be greater than or equal to 9.

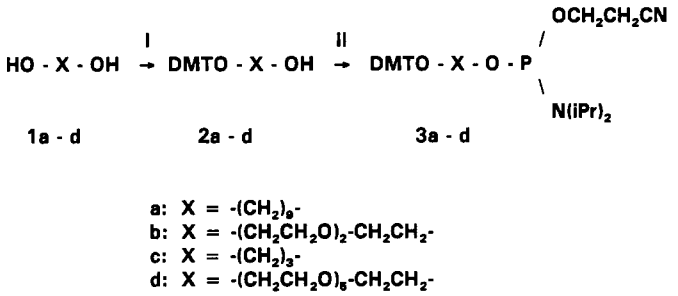


FIGURE 3: Chemical synthesis strategy for linker-derivatized phosphoramidites (**3a–d**) via dimethoxytritylation (step I) and phosphitylation (step II). Step I: DMT-Cl/pyridine. Step II: 2-Cyanoethyl *N,N*-diisopropylchlorophosphoramidite/DIPEA/anhydrous THF.

E), which should possess increased thermal stability (Ashley & Kushlan, 1991). Also, such closed duplexes should be resistant to 3′-exonucleases, which have been shown to be the primary source of oligonucleotide degradation in vivo (Shaw et al., 1991).

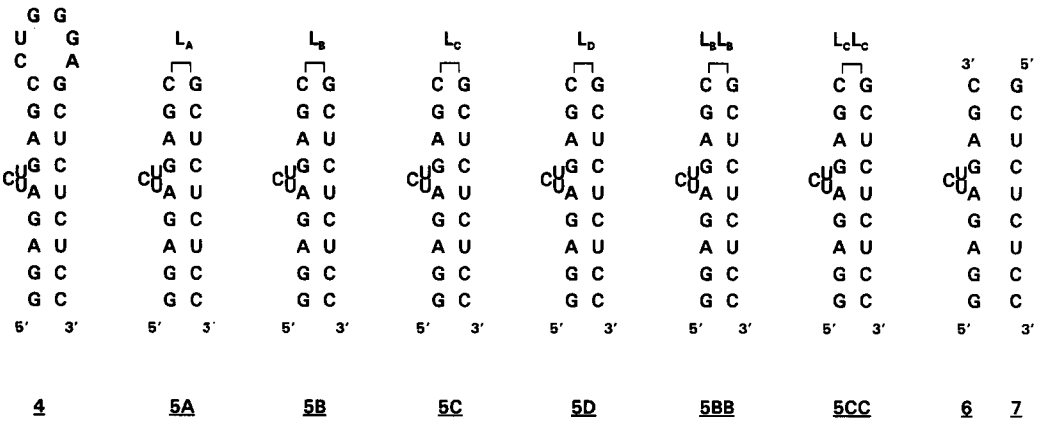
Several factors such as length, hydrophobicity, rigidity, steric effects, and electronic overlap are important in determining linker differences. In this study, the factors examined were length and hydrophobicity of the linker chain. Optimum base pairing in the duplex requires the heterocyclic bases involved in the formation of a Watson–Crick base pair to be positioned at a distance appropriate for hydrogen bond formation. If the synthetic linker is too short, the terminal base pair will be distorted or unable to form. If the linker is too long, the ends of the duplex will breathe apart and thus lose stability. Hydrophobicity was also examined because nucleotide backbones are highly charged. The incorporation of hydrophobic chains into synthetic oligonucleotides can facilitate their uptake across cell membranes (Shea et al., 1990; Saison-Behmoaras et al., 1991). On the basis of molecular modeling experiments (Figure 2), the minimum linker length between the terminal phosphates of a duplex is around 9 carbon atoms. At a shorter linker length a significant destabilizing effect would be generated. In contrast, the calculations showed no significant effect when the chain length was extended to 20 carbon atoms.

Four synthetic linkers were designed (Figure 3). Linker A (L_A, derived from 1,9-nonanediol) corresponds to the minimal length predicted by molecular modeling and is highly hydrophobic. Linker B (L_B, derived from triethylene glycol) has a length similar to that of L_A, except that it incorporates two oxygen atoms. Linker C (L_C, derived from 1,3-propanediol) has a three-carbon backbone and is the shortest linker in this series. L_C has the same number of carbon atoms as a single ribose backbone and is considered a "natural" substitute for any individual nucleotide, approximately retaining the correct internucleotide spacing. Linker D (L_D, derived from hexaethylene glycol) has 17-atom backbone and is the longest linker.

The chemistry employed in this work was chosen to be compatible with the standard phosphoramidite chemistry currently used for both DNA and RNA synthesis (Caruthers et al., 1987). The advantages of this synthetic approach are 3-fold: (1) the linker-derivatized phosphoramidites can be readily synthesized by a fast two-step procedure—dimethoxytritylation and phosphitylation (Figure 3); (2) they can be incorporated into oligonucleotides at any desired position, in an automated fashion, just like standard nucleoside phosphoramidites; and (3) it can be used to generate linkers of extended length which incorporate internal phosphate groups by multiple couplings of the linker phosphoramidites. This versatility has proven to be advantageous in the design of our model compounds.

Design and Synthesis of Linker-Derivatized Oligoribonucleotides. Most of the existing research on nucleic acid miniduplexes has dealt with their thermodynamic properties, such as equilibria between different structures in solution or thermal denaturation behavior. Few studies have examined the biological function of such derivatives. It is reasonable to expect that covalently linking two complementary single strands would increase dramatically the thermal stability of the resulting duplex. However, it does not necessarily follow that such a modified nucleic acid duplex would retain biological function(s). Inactivity could result from unpredictable distortions exerted by the synthetic linker or unfavorable interactions with the corresponding binding protein. It is critical to assay and assess the biological integrity of such modified duplex structures. The model system chosen in this work derives from the Tat–TAR complex of human immunodeficiency virus type 1 (Sumner-Smith et al., 1991). Previous work has demonstrated that the oligoribonucleotide TAR (Figure 4, oligomer **4**) can fold into a specific stem–bulge–loop structure and that it will bind in a specific fashion to both full-length Tat protein (86 amino acids) and a Tat-derived peptide (RKKRRQRRRPPQGS). The binding reaction can be easily followed by a RNA-based gel-retardation assay under nondenaturing conditions. Recent studies have also shown that the six-nucleotide loop in TAR is required for in vivo viral trans-activation, while it is dispensable for in vitro binding to Tat protein or Tat peptide (Roy et al., 1990b). These characteristics provided an ideal system for evaluating our approach, simply by replacing the 6-nt loop with synthetic linkers of different length and hydrophobicity. Six linker-derivatized TAR analogues were designed. The linkers varied in length from C_3 (**5C**) to C_{17} (**5D**) (note: this length does not include the two terminal phosphates) and varied from being a highly hydrophobic chain (**5A**) to "semi"-hydrophobic (**5B**), to highly charged (**5CC**). The effect of a negatively charged linker on duplex stability was examined by double coupling of the relevant phosphoramidite (**5BB** versus **5D**).

Chemical synthesis of these linker-derivatized TAR oligoribonucleotides was carried out under standard conditions for RNA synthesis except that the condensation reaction was increased to two sequential couplings of 6 min (Table I, step 3). The coupling yield ranged from 94 to 96% on the basis of trityl measurements. After standard deprotection and purification from 20% denaturing polyacrylamide gels, the



L_A: -OP(O)$_2$O-(CH$_2$)$_9$-OP(O)$_2$O-
L_B: -OP(O)$_2$O-(CH$_2$CH$_2$O)$_2$-CH$_2$CH$_2$-OP(O)$_2$O-
L_C: -OP(O)$_2$O-(CH$_2$)$_3$-OP(O)$_2$O-
L_D: -OP(O)$_2$O-(CH$_2$CH$_2$O)$_5$-CH$_2$CH$_2$-OP(O)$_2$O-
L_BL_B: -OP(O)$_2$O-(CH$_2$CH$_2$O)$_2$-CH$_2$CH$_2$-OP(O)$_2$O-(CH$_2$CH$_2$O)$_2$-CH$_2$CH$_2$-OP(O)$_2$O-
L_CL_C: -OP(O)$_2$O-(CH$_2$)$_3$-OP(O)$_2$O-(CH$_2$)$_3$-OP(O)$_2$O-

FIGURE 4: Design of the linker-derivatized oligoribonucleotides. The 6-nucleotide loop sequence in the model TAR RNA molecule (oligomer **4**) was replaced with synthetic linkers of different length (3 (**5C**) to 17 atoms (**5D**)) and hydrophobicity (**5A** and **5B**). The linker in oligomer **5CC** has a length similar to those of the oligomers **5A** and **5B** except it contains a negatively-charged internal phosphate. A similar comparison can be made between oligomers **5BB** and **5D**. The two linear oligomers **6** and **7** were used to form an unlinked duplex as a control.

Table I: Synthetic Cycle for the Preparation of Linker-Derivatized TAR Oligoribonucleotides

step	reagent or solvent	purpose	time (s)
1	dichloroacetic acid in CH_2Cl_2 (2.5:97.5; v/v)	detritylation	5 × 20
2	anhydrous CH_3CN	wash	90
3	activated phosphoramidites in anhydrous CH_3CN[a]	coupling	2 × 360
4	anhydrous CH_3CN	wash	20
5	HPLC grade CH_2Cl_2	wash	20
6	anhydrous CH_3CN	wash	20
7	DMAP/THF (6.5 g, 94 ml) Ac_2O/lutidine/THF (1:1:8; v/v/v)	capping	60
8	0.1 M I_2 in THF/lutidine/H_2O (160:40:4; v/v/v)	oxidation	60
9	anhydrous CH_3CN	wash	3 × 20

[a] The coupling reactions were carried out by premixing 0.5 M tetrazole with 0.15–0.30 M standard or modified phosphoramidites in anhydrous CH_3CN.

Table II: Summary of Thermal Stability (T_m), Binding Affinity (K_d), and Binding Capacity for the Duplex-Stabilizing Linkers[a]

oligomer	substitution	T_m (°C)	binding (K_d)	binding capacity (%)
4	6-nt loop (wt sequence)	60	+ (0.41)	45.9
5A	linker A/loop	61	+ (0.71)	40.1
5B	linker B/loop	58	+ (0.95)	42.6
5C	linker C/loop	56	–	
5D	linker D/loop	63	+ (0.66)	56.0
5BB	2 × linker B/loop	59	+ (1.13)	38.3
5CC	2 × linker C/loop	56	+ (0.43)	17.8
6 + 7	without connection	32	–	

[a] Linker A: -(CH$_2$)$_9$-. Linker B: -(CH$_2$CH$_2$O)$_2$-CH$_2$CH$_2$-. Linker C: -(CH$_2$)$_3$-. Linker D: -(CH$_2$CH$_2$O)$_5$-CH$_2$CH$_2$-. K_d values are expressed in nanomolar concentrations. +, strong binding; –, no binding.

purity of the modified TAR oligoribonucleotides was checked by 5′-end labeling with T4 polynucleotide kinase and [γ-^{32}P]-ATP and analytical PAGE. The purified RNA oligomers were characterized by comparing their migration rates against those of known RNA sequences of similar length, enzymatic RNA sequencing (Donis-Keller, 1980), and/or base-composition analysis (Seela & Kaiser, 1987).

Nucleic acids containing palindromic sequences can adopt two structures in solution, i.e., a monomeric hairpin or a dimeric bulged duplex (Xodo et al., 1986). When a nucleotide loop is used as the bridging chain, the thermodynamic equilibrium favors the monomeric hairpin form due to unfavorable mismatching effects presented in the bulged duplex (Xodo et al., 1991). When our linker–TAR analogues were used immediately after purification, some slower-moving bands, presumably bulge-duplexed dimers, were observed on native polyacrylamide gels (data not shown). The proportion of the dimer bands varied for each oligomer. If, however, the RNA oligomer was preannealed, most of the slower-moving dimer band shifted to the faster-moving monomer band. Therefore, all of the linker-derivatized RNA oligomers were preannealed under identical conditions prior to performing the binding assays.

Thermal Stability and Binding Properties of Linker-Derivatized RNA Miniduplexes. Durand et al. (1990) have reported thermodynamic studies of oligodeoxyribonucleotide duplexes containing a linker derived from a hexaethylene glycol chain. The melting temperature (T_m) of the hairpin with a pentathymidine (T_5) loop was 69 °C, while that of the hexaethylene glycol loop was 72 °C. A similar phenomenon was observed for the RNA analogues. A typical melting curve is illustrated in Figure 5, and the melting temperatures are summarized in Table II. The T_m values of the TAR analogues ranged from 56 to 63 °C, very close to that of the wild-type sequence (60 °C). This result supports previous suppositions that there is no stabilizing interaction in TAR between the 6-nt loop and the RNA stem and that, consequently, the loop is available for interaction with cellular factors (Gaynor et al., 1989; Sheline et al., 1991; Rounseville & Kumar, 1992).

Thermal denaturation experiments indicated that every linker-derivatized TAR analogue had some secondary structure. However, this type of thermodynamic assessment could not verify whether or not the secondary structure retained the biologically active TAR stem–bulge structure normally assured

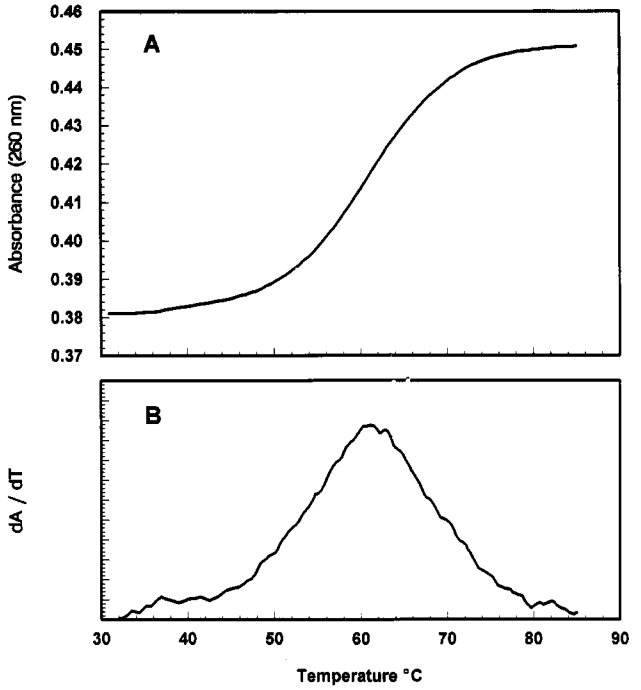


FIGURE 5: (Panel A) Thermal denaturation profile of oligomer **5A** in 100 mM NaCl/10 mM sodium phosphate (pH 7.0) at an oligomer concentration of 3 μM. The samples were heated from 25 to 85 °C in 1 °C increments, and the absorbance was monitored at 260 nm. (Panel B) First derivative of the melting curve measured in panel A and determination of the melting temperature (T_m).

by the 6-nt loop. Previous work on the Tat–TAR interaction has demonstrated that binding of Tat protein or Tat-derived peptide is very sensitive to subtle structural changes in the TAR RNA stem–loop (Summer-Smith et al., 1991). It is possible to assess the impact of a synthetic linker at one end of the TAR RNA by its ability to bind to Tat peptide. In the presence of a 500-fold excess of Tat-derived peptide, all of the linker–TAR oligomers showed binding, except for the oligomer containing only L_C (**5C**) (Figure 6).

In order to assess possible differences in binding affinity for the short Tat peptide and full-length native Tat protein, the binding affinities of the linker-derivatized TAR analogues for full-length Tat protein (86 amino acids) were assessed. Figure 7 shows a typical binding assay result, using a constant amount of oligomer **5A** and an increasing amount of full-length Tat protein. The K_d value for the full-length Tat (1.17 nM) was slightly higher than that for the Tat-derived peptide (0.71 nM). When TAR analogue **5B** was added to a preformed complex between the 27-nt fragment of the wild-type TAR stem-loop and full-length Tat protein, strong competition with the TAR sequence was observed (Figure 8). The complex was totally competed away when the ratio between the linker TAR analogue and the Tat protein was 1:1.

The initial qualitative Tat binding assays showed that five of the six linker-derivatized TAR analogues were able to bind to Tat-derived peptide. Quantitative binding assays were performed to obtain a more precise comparison between the TAR analogues. Briefly, starting with a fixed amount of ^{32}P-labeled TAR RNA analogue, the binding reactions were followed by adding an increasing amount of unlabeled Tat-derived peptide over a wide range of concentrations. The amount of material in each spot was quantified and the data were fit to the equation for a simple bimolecular equilibrium. Two useful parameters emerge from this set of experiments, the equilibrium binding constant (K_d) and the RNA binding capacity (Table II). The K_d values reflect the concentration at which 50% of "active RNA" forms a RNA–peptide complex. The binding capacity indicates the percentage of active RNA molecules which were capable of binding to peptide upon saturation. Only a fraction of the input RNA molecules (40–60%) are "biologically active"; the rest are considered to be "biologically inactive" as a result of probable misfolding into nonproductive conformations. The binding capacity does not affect the K_d value, but it is important for comparison of the linkers since only the correctly folded RNA molecules can be used for pharmaceutical purposes. As a consequence, assessment of different linkers was accomplished by comparing three different parameters: the thermal stability as measured by T_m, the binding affinity as measured by K_d, and the biological activity as measured by binding capacity.

DISCUSSION

Nucleic acid miniduplexes have the advantage of being able to bind proteins or peptides, yet they are short enough to be pharmaceutical agents. One goal of this work was to define the necessary characteristics of synthetic linkers which can stabilize such a nucleic acid miniduplex without interfering with its biological function. A "preferred length" for duplex-stabilizing linkers was predicted using the specific spatial requirements of a Watson–Crick base pair. Surprisingly, as demonstrated by RNA–peptide binding assays and thermal denaturation experiments (Table II), this preferred length covers a broad size range. Only slight differences in binding affinity and/or T_m were observed between L_A or L_B (9 internal atoms) and L_D (17 internal atoms) in the model system. By contrast, there is a definite lower length limit below which the resulting duplex structures were biologically inactive (e.g., oligomer **5C**, three internal atoms). These results support our expectation that a minimum linker length is required. The T_m of oligomer **5C** is only a few degrees lower than that of the other analogues and has a similar R_f on native gels when it is preannealed. This suggests that although the **5C** derivative folds as a monomeric hairpin structure (as opposed to a dimeric bulged-duplex structure), it must utilize the terminal CG base pair of the stem as part of the turn (a putative 5′-C-L_C-G-3′ loop). This structural change totally abolishes its peptide binding ability.

No benefit was observed when a negative charge was added into the duplex-stabilizing linker. Indeed, a neutral bridging linker appeared to have some advantages over a negatively charged linker. Oligomers **5D** and **5BB** incorporated linkers of similar length and composition, except that **5BB**'s linker contained an internal phosphate. Interestingly, the latter oligomer exhibited a lower T_m, a higher K_d, and a lower binding capacity. A similar result was observed on comparison of the "neutral" analogues **5A** and **5B** versus the "charged" analogue **5CC**. The situation in this case was complicated by the fact that oligomer **5CC** had a lower K_d value. However, this derivative showed very little binding to Tat-derived peptide (with a binding capacity of only 18% as compared to 40% for oligomers 5A and 5B), indicating a high proportion of this analogue folded into an alternative structure(s).

Thermal denaturation experiments indicated that a remarkable increase in thermal stability was conferred by connecting the two ends of a short RNA duplex. The increases in T_m were 24–31 °C higher than that of the unlinked duplex (oligomer **6** + oligomer **7**). All of the linker-derivatized TAR analogues showed T_m's well above 37 °C, suggesting that they would remain stable as double-stranded duplexes under physiological conditions. However, no correlation between T_m and biological activity, as measured by peptide binding in the gel shift assay, was observed.

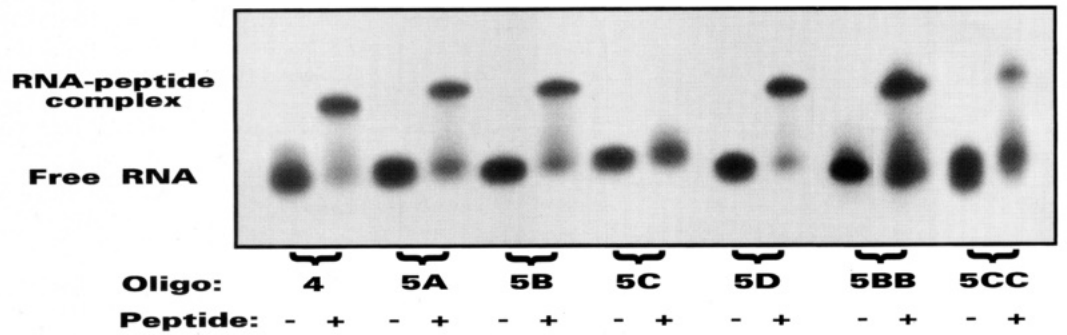


FIGURE 6: RNA gel-shift binding assays of the linker-derivatized TAR analogues. The 5′-^{32}P-end labeled oligomers **5A**–**5CC** were incubated in the absence (–) or presence (+) of Tat-derived peptide. The peptide–RNA complexes were analyzed on a 5% nondenaturing polyacrylamide gel as described in Materials and Methods. Oligomer **4** was used as a positive control. Only the oligomer derived from the shortest linker (L_C, oligomer **5C**) failed to form any complexes with Tat-derived peptide.

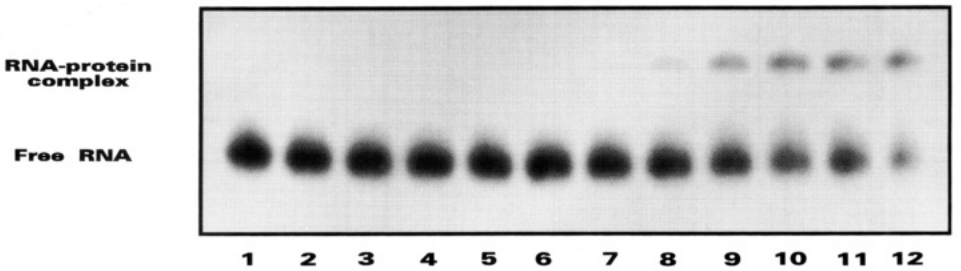


FIGURE 7: RNA gel-shift binding assay of oligomer **5A** with the full-length HIV-1 Tat-protein. ^{32}P-Labeled oligomer **5A** (1×10^{-10} M) was incubated at room temperature with increasing amounts of Tat protein. Tat protein concentrations were as follows: lane 1, 0 M; lane 2, 1×10^{-11} M; lane 3, 5×10^{-11} M; lane 4, 1×10^{-10} M; lane 5, 5×10^{-10} M; lane 6, 1×10^{-9} M; lane 7, 5×10^{-9} M; lane 8, 1×10^{-8} M; lane 9, 5×10^{-8} M; lane 10, 1×10^{-7} M; lane 11, 5×10^{-7} M; lane 12, 1×10^{-6} M. The complexes were analyzed on nondenaturing polyacrylamide gels at 4 °C. The complex began to appear when the ratio between the oligomer and the Tat protein reached ~50:1 (lane 7).

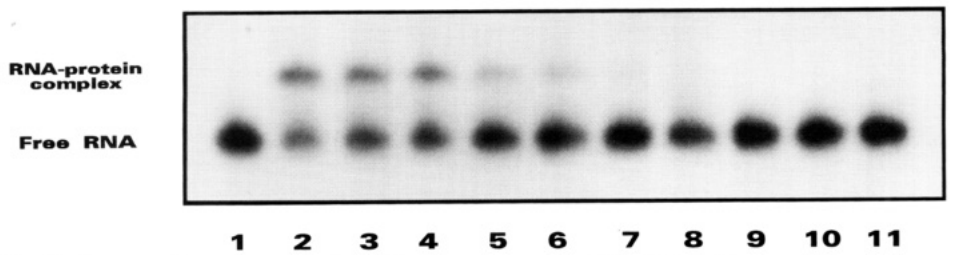


FIGURE 8: Competition binding experiment between the wild-type sequence **4** and the linker oligomer **5B** in the presence of the full-length Tat protein. ^{32}P-Labeled oligomer (1×10^{-10} M) **4** was premixed with 1×10^{-7} M Tat protein. An increasing amount of the linker oligomer **5B** was added. Lane 1, ^{32}P-labeled oligomer **4** alone; lane 2, complex formation between the Tat protein and the wild-type oligomer **4** in the absence of the linker oligomer **5B**; lanes 3–11, oligomer **5B** was added as a competitor at the following concentrations—lane 3, 5×10^{-10} M; lane 4, 1×10^{-9} M; lane 5, 5×10^{-9} M; lane 6, 1×10^{-8} M; lane 7, 5×10^{-8} M; lane 8, 1×10^{-7} M; lane 9, 5×10^{-7} M; lane 10, 1×10^{-6} M; lane 11, 5×10^{-6} M.

These results are encouraging, and a number of additional linker analogues are currently under investigation. It should be possible to couple these miniduplexes to an affinity column, reporter group, or a reactive moity via a tether bridge using a modified synthetic scheme (e.g., H-phosphonate approach) (Figure 1, structure F). This would introduce a neutral internal phosphotriester linkage rather than a destabilizing negatively-charged phosphodiester linkage.

In conclusion, we have demonstrated, by chemical synthesis and RNA gel-shift binding assays, that synthetic linkers of specific design can be used to generate and stabilize a specific nucleic acid duplex structure. The modified TAR RNA analogues synthesized in this study have similar thermal stability as the wild-type sequence where the strand connection is made by a 6-nt loop. Most importantly, these RNA miniduplexes of only 9-bp retain their full binding capacity to full-length Tat protein and Tat-derived peptide. These results clearly demonstrate the general utility of nucleic acid miniduplexes as a new class of nucleic acid analogue with applications in both fundamental research and pharmaceutical development.

REFERENCES

Andrus, A. (1988) *Applied Biosystems, User Bulletin No. 47.*

Antao, V. P., Lai, S. Y., & Tinoco, I., Jr. (1991) *Nucleic Acids Res. 19*, 5901–5905.

Ashley, G. M., & Kushlan, D. M. (1991) *Biochemistry 30*, 2927–2933.

Bielinska, A., Shivdasani, R. A., Zhang, L., & Nabel, G. J. (1990) *Science 250*, 997–1000.

Caruthers, M. H., Barone, A. D., Beaucage, S. L., Dodds, D. R., Fisher, E. F., McBride, L. J., Matteucci, M., Stabinsky, Z., & Tang, J.-Y. (1987) *Methods Enzymol. 154*, 287–313.

Cowart, M., & Benkovic, S. J. (1991) *Biochemistry 30*, 778–796.

Delling, U., Roy, S., Sumner-Smith, M., Barnett, R., Reid, L., Rosen, C. A., & Sonenberg, N. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 6234–6238.

Donis-Keller, H. (1980) *Nucleic Acids Res. 8*, 3133–3142.

Durand, M., Chevrie, K., Chassignol, M., Thuong, N. T., & Maurizot, J. C. (1990) *Nucleic Acids Res. 18*, 6353–6359.

Ferentz, A. E., & Verdine, G. L. (1991) *J. Am. Chem. Soc. 113*, 4000–4002.

Gaynor, R., Soultanakis, E., Kuwabara, M., Gania, J., & Sigman, D. S. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4858–4862.
Harel-Bellan, A., Brini, A., Ferris, D. F., Robin, P., & Farrar, W. L. (1989) *Nucleic Acids Res. 17*, 4077–4087.
Helene, C., & Toulme, J.-J. (1990) *Biochim. Biophys. Acta 1049*, 99–125.
Itakura, K., Rossi, J. J., & Wallace, R. B. (1984) *Annu. Rev. Biochem. 53*, 323–356.
Matteucci, M., & Webb, T. (1987) *Tetrahedron Lett. 28*, 2469–2472.
Ogilvie, K. K., Usman, N., Nicoghosian, K., & Cedergren, R. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5764–5768.
Petric, A., Bhat, B., Leonard, N. J., & Gumport, R. I. (1991) *Nucleic Acids Res. 19*, 585–590.
Rounseville, M. P., & Kumar, A. (1992) *J. Virol. 66*, 1688–1694.
Roy, S., Delling, U., Chen, C.-H., Rosen, C. A., & Sonenberg, N. (1990a) *Genes Dev. 4*, 1365–1373.
Roy, S., Parkin, N. T., Rosen, C. A., Itovitch, J., & Sonenberg, N. (1990b) *J. Virol. 64*, 1402–1406.
Saison-Behmoaras, T., Tocque, B., Rey, I., Chassignol, M., Thuong, N. T., & Helene, C. (1991) *EMBO J. 10*, 1111–1118.
Scaringe, S. A., Franckyl, C., & Usman, N. (1990) *Nucleic Acids Res. 18*, 5433–5441.
Seela, F., & Kaiser, K. (1987) *Nucleic Acids Res. 15*, 3113–3129.
Shaw, J.-P., Kent, K., Bird, J., Fishback, J., & Froehler, B. (1991) *Nucleic Acids Res. 19*, 747–750.
Shea, G. R., Marsters, J. C., & Bischofberger, N. (1990) *Nucleic Acids Res. 18*, 3777–3783.
Sheline, C. T., Miloco, L. W., & Jones, K. A. (1991) *Genes Dev. 5*, 2508–2520.
Sumner-Smith, M., Roy, S., Barnett, R., Reid, L. S. Kuperman, R., Delling, U., & Sonenberg, N. (1991) *J. Virol. 65*, 5196–5202.
Uhlmann, E., & Peyman, A. (1990) *Chem. Rev. 90*, 543–584.
Weeks, K., Ampe, C., Schultz, S. C., Steitz, T. A., & Crothers, D. M. (1990) *Science 249*, 1281–1285.
Xodo, L. E., Manzini, G., Quadrifoglio, F., van der Marel, G. A., & van Boom, J. H. (1986) *Nucleic Acids Res. 13*, 5389–5398.
Xodo, L. E., Manzini, G., Quadrifoglio, F., van der Marel, G. A., & van Boom, J. H. (1991) *Nucleic Acids Res. 19*, 1505–1511.