Nucleic Acids Research, 1998, Vol. 26, No. 22 **5093–5101**

Construction of a directed hammerhead ribozyme library: towards the identification of optimal target sites for antisense-mediated gene inhibition

Michael L. Pierce[1] and Duane E. Ruffner*

Department of Pharmaceutics and Pharmaceutical Chemistry and [1]Department of Bioengineering, University of Utah, 421 Wakara Way, Suite 318, Salt Lake City, UT 84108, USA

Received August 11, 1998; Revised September 9, 1998; Accepted September 24, 1998

ABSTRACT

Antisense-mediated gene inhibition uses short complementary DNA or RNA oligonucleotides to block expression of any mRNA of interest. A key parameter in the success or failure of an antisense therapy is the identification of a suitable target site on the chosen mRNA. Ultimately, the accessibility of the target to the antisense agent determines target suitability. Since accessibility is a function of many complex factors, it is currently beyond our ability to predict. Consequently, identification of the most effective target(s) requires examination of every site. Towards this goal, we describe a method to construct directed ribozyme libraries against any chosen mRNA. The library contains nearly equal amounts of ribozymes targeting every site on the chosen transcript and the library only contains ribozymes capable of binding to that transcript. Expression of the ribozyme library in cultured cells should allow identification of optimal target sites under natural conditions, subject to the complexities of a fully functional cell. Optimal target sites identified in this manner should be the most effective sites for therapeutic intervention.

INTRODUCTION

Short DNA and RNA oligonucleotides and ribozymes can be designed to bind specifically to any chosen mRNA. Through a variety of mechanisms, this binding can specifically block expression of the mRNA. Consequently, this so-called antisense-mediated gene inhibition has the potential to be used therapeutically for the treatment of disease resulting from aberrant gene expression (1–3, reviewed in 4).

One major parameter determining efficacy of any antisense strategy is target site accessibility on the chosen transcript. Since the antisense agent needs to base pair with the mRNA in order to elicit its effect, areas of the transcript not involved in extensive secondary structure should be better targets than those that are highly structured. Towards this, a number of researchers have used computer algorithms to predict the structure of the mRNA target, subsequently selecting target sites void of secondary structure. This type of approach has met with mixed success (5–8). While the level of accuracy of the computer algorithms used to fold the RNA could be called into question, another explanation for the varied success could be that the programs do not predict higher order structure, which can be an important determinant of target site accessibility (9). Finally, an antisense agent that folds into a highly stable structure of its own may be unable to interact with an otherwise accessible target site.

In addition to structure, other factors likely influence target accessibility. Messenger RNA is associated with heterogeneous nuclear ribonucleoproteins in the nucleus, which on transport to the cytoplasm are replaced by cytoplasmic ribonucleoproteins (reviewed in 10). Bound ribonucleoproteins are likely to occlude potential target sites and/or influence the folding of the mRNA. As a consequence of the unknown effects of ribonucleoproteins, the value of structure prediction in target site selection is questionable.

With respect to antisense ribozymes, especially the hammerhead, additional factors can influence activity. The sequence of the target mRNA surrounding the cleavage site can alter the cleavage rate (11–13) as well as the turnover rate of the ribozyme (14). Cleavage activity can be enhanced or diminished by a variety of facilitators, including proteins and small molecules (15–20). The effects of facilitators may vary for ribozymes targeting different sites, due to sequence differences of the ribozymes. These influences can affect the suitability of a target site, independent of the accessibility of the site.

Clearly, optimum target site selection involves many considerations and is beyond our current ability to predict. Consequently, empirical methods represent the only way to identify the most effective site(s) and this requires examining every potential target site. Towards this goal several groups have used non-directed approaches to target site selection (6,21–25). In each of these studies, degenerate oligonucleotides were used in cell-free screens. This presents two problems. First, the cell-free screens have produced incomplete correlations between *in vitro* and *in vivo* activity and, therefore, their effectiveness in identifying *in vivo* therapeutic targets has yet to be demonstrated (reviewed in 26). Second, the use of degenerate libraries necessitates searching through a prohibitive number of sequences (2.68×10^8 molecules for hybridizing regions totaling 14 nt). Furthermore, degenerate

*To whom correspondence should be addressed. Tel: +1 801 585 5093; Fax: +1 801 585 3614; Email: duane.ruffner@m.cc.utah.edu

libraries target not only the transcript of interest, but all cellular RNAs. In this regard, they have the ability to block expression of all mRNAs, as well as inhibit the activity of structural RNAs such as rRNA and tRNA. This prevents the use of degenerate libraries in cell-based screens.

Others have examined a directed approach to target site selection (27). In this case, a library of cDNA fragments specific to the mRNA of interest was used. This overcomes some of the disadvantages of the non-directed approaches. However, since this screen was performed in cell lysates, it is not clear that the targets identified are suitable for *in vivo* targeting. This is supported by the incomplete correlation between the *in vitro* and *in vivo* activities of the identified sites.

Here we describe an alternative directed approach. This approach uses 'directed' hammerhead ribozyme libraries. The hammerhead ribozyme was chosen since, due to its cleavage activity, identification of effective target sites should be based primarily on target site accessibility. Consequently, identified sites should also be suitable targets for antisense oligodeoxynucleotides that operate by a RNase H-dependent mechanism. Additionally, since the library targets all positions, not just those that contain the NUH sequence required for hammerhead-mediated cleavage (11,12,28,29) (where H represents A, C or U as defined by the International Union of Biochemistry; 30), target sites that can give rise to gene inhibition through non-cleavage-dependent mechanisms can also be identified. Since the ribozyme library is directed, much smaller libraries can be used and effects on non-target genes are eliminated. This allows for the possibility that these libraries can be expressed and assayed in intact cells, where optimal target sites can be identified under the conditions in which therapeutic antisense molecules are expected to operate. Towards this, we describe the successful construction of a ribozyme library targeted to the essential transcriptional activator, ICP4, of the Herpes simplex virus (HSV).

MATERIALS AND METHODS

Materials

*Bsp*HI, *Sap*I, *Hph*I, *Bsm*FI, *Xba*I, *Pst*I, T4 DNA ligase, Vent polymerase (exo$^+$), exonuclease III (ExoIII), T4 DNA polymerase, Klenow fragment of DNA polymerase I and mung bean nuclease were all purchased from New England Biolabs (Beverly, MA). Pfu polymerase was purchased from Stratagene (La Jolla, CA). Deoxyribonucleotides were purchased from Pharmacia (Piscataway, NJ). Microcon 50 spin concentrators were purchased from Amicon (Beverly, MA). Shrimp alkaline phosphatase and Sequenase v.2.0 sequencing kit were purchased from Amersham (Arlington Heights, IL). The plasmids pC194 [containing the coding region for chloramphenicol acetyltransferase (CAT)] and EBOpLPP were obtained from ATCC (Rockville, MD), and the plasmid pRC/CMV was purchased from Invitrogen (Carlsbad, CA). The plasmid containing the HSV-1 ICP4 genomic fragment, pTEG2 (31), was a gift from Saul Silverstein. Oligonucleotides were synthesized on an Applied Biosystems automated oligonucleotide synthesizer at the University of Utah oligonucleotide/peptide core facility.

Construction of pRbzlib

The *Hin*dIII–*Hpa*I fragment of pLA2917 (32), containing the kanamycin resistance gene, was inserted into *Hin*dIII/*Sma*I-digested pUC19 to produce pUCKan. An *Hph*I and two *Bsa*HI sites were eliminated from the kanamycin resistance gene by site-directed mutagenesis to produce pUCKan*. The mutagenized kanamycin resistance gene was removed by *Hin*dIII/*Eco*RI digestion, the termini were blunted by 5′-overhang fill-in using the Klenow fragment and ligated to the blunted 843 bp *Bsp*HI–*Sap*I fragment of pUC19 containing the origin of replication. A clone (pKan) was selected in which the *Eco*RI and *Bsp*HI sites were juxtaposed. The *Bsm*FI and *Pst*I sites were eliminated from pKan by site-directed mutagenesis using the procedure of Merino *et al.* (33). The multiple cloning site for pRbzlib was constructed from the overlapping oligodeoxynucleotides MCS-L (5′-AAGCT TGGTG ACTGT CTTCG AGCTC GAATT CATCG ATATC TAGAG TTTA-3′) and MCS-R (5′-GTCGA CGGGA CTGCA GGTTT AAACT CTAGA TATC-3′) by 5′-overhang fill-in with the Klenow fragment of DNA polymerase I. The double-stranded multiple cloning site was inserted into *Eco*RI-linearized and blunted pKan to make pRbzlib.

Construction of pRbzshuttle

A hygromycin expression cassette capable of being expressed in both mammalian and prokaryotic systems was constructed using overlap extension PCR. The 1026 bp hygromycin resistance coding sequence from EBOpLPP was joined at its 3′-end to the 322 bp 3′-untranslated region (UTR)/SV40 early polyadenylation sequence from pRC/CMV, while the 527 bp dual ampicllin/SV40 early promoter from pEGFP-1 was joined to the 5′-end. The sequence of the primers used in the PCR are: 3′-UTR/poly(A) segment, CCGAG GGCAA AGGAA TAGGC GGGAC TCTGG GGT and CTCGA GGTCG ACGGG ATCCA G; hygromycin coding region, GGATG AGGAT CGTTT CGCAT GAAAA AGCCT GAA and ACCCC AGAGT CCCGC CTATT CCTTT GCCCT CGG; amp/SV40 early promoter, CGTCA GGTGG CACTT TTCGG and TTCAG GCTTT TTCAT GCGAA ACGAT CCTCA TCC. Each portion of the hygromycin cassette was prepared by PCR using one of the three primer sets and the appropriate template. The resulting fragments were gel purified. The hygromycin encoding and the 3′-UTR/poly(A) fragments were combined and used in a second PCR reaction to produce the hygromycin–3′-UTR/poly(A) fragment. In a final PCR, this fragment was combined with the amp/SV40 fragment to produce the complete 1875 bp hygromycin gene cassette. The hygromycin gene cassette was ligated into the 843 bp *Bsp*HI–*Sap*I oriP-containing fragment of pUC19, producing pHyg. The 4914 bp *Eco*RI–*Bam*HI fragment containing the EBNA-1 and EBV oriP sequences from EBOpLPP was inserted between the hygromycin cassette and the pUC19 origin of *Xho*I-digested pHyg to make pEBV. The 1060 bp expression cassette was excised from pRC/CMV using *Nru*I and *Pvu*II and inserted into the *Bam*HI site of pEBV to produce pRbzshuttle.

Construction of a hammerhead ribozyme catalytic core cassette

A cassette encoding the hammerhead catalytic core, interrupted by the CAT gene (34), was constructed using PCR as follows. PCR primers were prepared that are complementary to the CAT gene on their 3′-ends and encode the hammerhead catalytic core on their 5′-ends. The sequences of the primers were as follows: CatCass 1, 5′-CTGATGAGGTCG*cga*ctagtgttgacaat-3′; CatCass 2, 5′-TTCGGTCTCG*cga*gcaggttagtgaca-3′. Upper case letters

encode the ribozyme core sequence. Non-italicized lower case letters encode sequences on the termini of the CAT gene. Underlined letters indicate the position of *Nru*I restriction sites. The PCR reaction contained 5 ng CAT gene DNA, 100 pmol each CatCass 1 and CatCass 2, 1 mM each dNTP, 5 U Vent polymerase (exo^+) in the standard Vent polymerase buffer with the exception that the concentration of $MgSO_4$ was increased to 5.2 mM. The use of Vent polymerase ensures that the cassette possesses blunt ends. The mixture was incubated at 94°C for 2 min and cycled as follows: 94°C 1 min, 45°C 30 s, 72°C 2 min, for 5 cycles; followed by 94°C 30 s, 60°C 15 s, 72°C 2 min, for 15 cycles; followed by 73°C for 5 min. The cassette was separated from unincorporated primers by agarose gel electrophoresis and recovered using standard procedures.

Construction of the anti-ICP4 ribozyme library

The library was constructed as illustrated in Figure 3. The optimized conditions for construction of the library were as follows. First, the 4489 bp *Bgl*II–*Eco*RI fragment from pTEG2 (31) was cloned into *Eco*RI/*Eco*RV-digested pRbzlib. This fragment includes 125 bp upstream of the translational start site, 466 bp downstream of the translation termination sequence and the entire genomic coding sequence of ICP4. The resulting clone, pRbzlib-ICP4, possesses the ICP4 fragment with the sense strand as the upper strand. From pRbzlib-ICP4 a deletion library was produced as follows. Twenty micrograms of CsCl gradient purified plasmid DNA was digested with *Pst*I and *Xba*I and subsequently concentrated and desalted using a microcon 50 spin filter. The DNA was brought to a volume of 60.4 µl in ExoIII buffer, warmed to 37°C, and 300 U of ExoIII were added. At 1 min intervals after the addition of the ExoIII, 2.5 µl of the reaction were removed and placed into microfuge tubes on ice, containing 7.5 µl of mung bean nuclease buffer (66.7 mM NaOAc, pH 5.2, 200 mM NaCl, 1.3 mM $ZnCl_2$ and 1 U of mung bean nuclease). After 25 time points had been taken, the tubes were placed at 20°C for 30 min. After 30 min, all tubes were combined and extracted once each with phenol:$CHCl_3$ (50:50) and $CHCl_3$ and precipitated with 2 vol 100% ethanol. The dried DNA pellet was suspended in 18 µl of deionized H_2O and 2.5 µl of 10 mM each dNTP, 2.5 µl 10× Pfu polymerase buffer and 5 U of Pfu polymerase were added. The mixture was incubated at 72°C for 15 min. The plasmid DNA was circularized by ligation for 4 h at room temperature, in a 1.25 ml reaction containing 5% PEG. Except for the modifications indicated, all ligations were performed with T4 DNA ligase under the conditions suggested by the manufacturer. The ligation mixture was transformed into DH5α and grown in liquid culture. The resulting deletion library was used to produce the 14 bp fragment library as follows. Two micrograms of the deletion library were digested with *Bsm*FI and *Bbs*I and the ends blunted with Pfu polymerase as before. After gel purification, the DNA was circularized by ligation for 4 h at room temperature, in a 600 µl reaction containing 5% PEG. The ligation mixture was transformed into DH5α and grown in liquid culture. The resulting 14 bp fragment library was used to prepare the ribozyme library as follows. One microgram of the fragment library was digested with 8 U of *Hph*I for 1 h at 37°C and the ends were polished with T4 DNA polymerase. The hammerhead cassette was inserted by ligating 500 ng of the *Hph*I-digested library with 5 µg of the ribozyme core cassette. The ligation product was transformed into DH5α and grown in culture under chloramphenicol selection. After purification, 2 µg of the plasmid were digested with *Hin*dIII and *Sal*I and the terminal phosphates were removed using shrimp alkaline phosphatase. The *Hin*dIII/*Sal*I digest was fractionated on an agarose gel and the dephosphorylated ribozyme/chloramphenicol cassette was recovered using standard procedures. The cassette was combined with an equimolar amount of *Hin*dIII/*Xho*I-digested pRbzshuttle and ligated using a modified two-step procedure (35) (the first step was performed at room temperature for 1 h and the second at 16°C overnight). The ligation was transformed into DH5α and grown in culture under chloramphenicol selection. One microgram of plasmid DNA was purified and digested with *Nru*I to release the chloramphenicol gene and recircularized by ligation in a volume of 600 µl. The final ligation product was transformed into DH5α and plasmid DNA recovered and purified on a CsCl gradient.

RESULTS AND DISCUSSION

The hammerhead ribozyme can be separated into two domains, a catalytic domain composed of most of the conserved core and helix–loop II and a substrate recognition domain (Fig. 1). Conceptually, a gene encoding a hammerhead ribozyme can be produced in three steps. First, a short (10–20 bp) double-stranded DNA fragment encoding the sequence of the desired target and containing the required NUH sequence is bisected. Second, a single base pair corresponding to the nucleotide 5′ of the cleavage site (H_{17}) is deleted. Finally, a fragment encoding the catalytic core is inserted in place of the deleted base pair. We use just such a process to prepare not one, but an entire library of ribozymes targeting all sites on the essential transcriptional activator, ICP4, of the HSV. While this library contains ribozymes targeting all cleavable sites on the target, all non-cleavable sites are also targeted.

ICP4 was chosen for two reasons. First, since it is the major regulatory protein of HSV-1 and is essential for viral replication (31,36–38), its inhibition could be used for the therapeutic treatment of HSV infection. Second, because it is a transcriptional activator, inhibition of ICP4 expression can be measured indirectly by measuring the expression of a reporter gene that is driven by an ICP4-dependent promoter. This allows target site accessibility and antisense efficacy to be determined on the native transcript. This is important since any alteration, such as fusion with a reporter sequence, could alter the pattern of accessibility. Consequently, accessible sites identified on an altered transcript may have little or no therapeutic value.

To allow construction and expression of an anti-ICP4 ribozyme library, two plasmid vectors were constructed. The plasmid pRbzlib was designed to allow production of ribozyme libraries against any mRNA, ICP4 in our case, and pRbzshuttle allows expression of that library in mammalian cells (Fig. 2).

Vector and catalytic core design considerations

pRbzlib possesses the pUC19 origin of replication and a kanamycin resistance gene allowing selection in bacterial cells (Fig. 2A). The kanamycin resistance gene was chosen as the selectable marker since, of all the available bacterial selection markers, it possessed the fewest sites that are present in our multiple cloning site (MCS). Therefore, it was the simplest to modify site specifically to eliminate undesirable sites. The MCS was carefully engineered and possesses the following salient features. It possesses a short polylinker that allows much

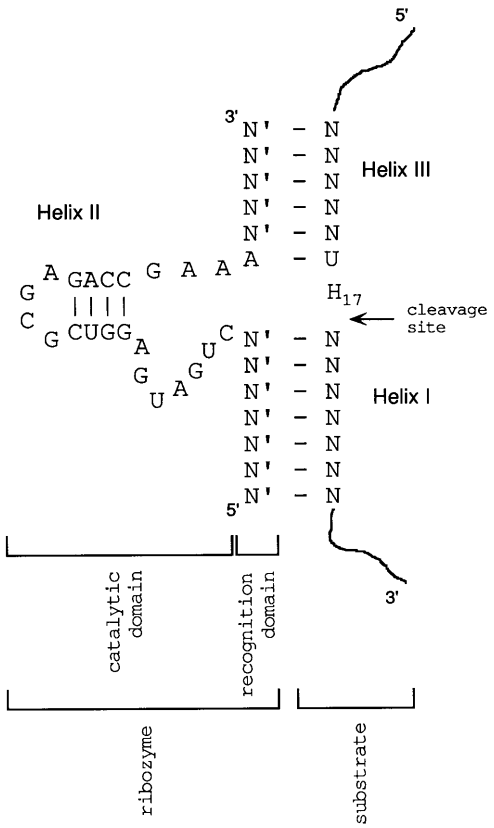

Figure 1. Secondary structure of the hammerhead ribozyme associated with substrate. H at position 17 designates A, C or U as defined by the IUB (30).

flexibility in the cloning of the gene or cDNA sequence of interest, which represents the first step in construction of a ribozyme library. The polylinker includes several restriction sites that leave sticky ends. These sites can be used to directionally clone the cDNA or genomic fragment in the correct orientation. Alternatively, the fragment can be cloned by blunt-end ligation and the correctly oriented clone can be selected by restriction analysis. The *Pst*I and *Pme*I sites allow the generation of a substrate for unidirectional digestion by ExoIII into the cloned cDNA or genomic fragment. This allows preparation of a serial deletion library of the cloned insert. The *Bsm*FI and *Bbs*I sites are used together to convert the deletion library into a 14 bp fragment library. The *Hph*I site allows bisection of the 14 bp fragment library for introduction of the hammerhead catalytic core.

The introduction of the catalytic core presents several difficulties. The core must be inserted by blunt-end ligation and in the correct orientation to produce a functional ribozyme. Additionally, due to its small size (Fig. 1), it is difficult to prevent the introduction of concatamers of the core and/or contamination of the library with clones that do not acquire a catalytic core. To increase the effectiveness and efficiency of this step we designed a catalytic core that was interrupted by the CAT gene (Fig. 3). CAT selection allows use of a non-phosphorylated cassette. This prevents insertion of multimers and selects against non-recombinants. Additionally, the CAT gene allows selection of clones acquiring a correctly oriented catalytic core. In the desired orientation, transcription of the CAT and kanamycin genes is in the same direction. In the incorrect orientation, CAT expression is inhibited by antisense expression from the kanamycin resistance gene. This effect has been reported elsewhere (39). After selection, the CAT gene is removed by digestion with *Nru*I to produce a sequence encoding a hammerhead ribozyme.

pRbzshuttle (Fig. 2B) was designed to allow replication and expression of the ribozyme library in mammalian cells. It possesses a MCS for insertion of the ribozyme library. The MCS is flanked on one end by a dual CMV/T7 promoter allowing expression of the ribozyme gene both in mammalian cells as well as by *in vitro* transcription using T7 RNA polymerase. On the other end of the MCS is a bovine growth hormone polyadenylation signal for efficient expression in mammalian cells. pRbzshuttle possesses a hygromycin resistance gene driven by a dual promoter to allow selection in bacterial and mammalian cells. The pUC19 origin of replication allows replication in bacterial cells. For replication in mammalian cells the EBV origin and the EBNA-1 gene were included (40,41).

The EBV origin and EBNA-1 gene are essential for identification of effective antisense targets in mammalian cells. These sequences allow extrachromosomal replication of pRbzshuttle. Episomal expression is important for several reasons. First, it eliminates the clone-to-clone variation in expression that occurs if stable transfectants are used (42). Second, since the copy number of the episomal vector is determined primarily by the transfection conditions and once established remains tightly regulated (43), then effects on expression due to differences in copy number should be minimal. Consequently, our selection of ribozyme efficacy should be based on accessibility and not the level of expression. Third, the use of an episomal expression vector allows for high transfection efficiency (42,44). This is important to ensure that all ribozymes present in our library are represented in the mammalian transfectants. Finally, the plasmid can be recovered and shuttled back into bacterial cells. This allows the sequence of effective ribozymes to be determined, thereby identifying accessible target sites.

To demonstrate episomal replication, pRbzshuttle was used to transfect HeLa cells and the cells were grown in culture under 400 μg/ml hygromycin selection. After 1 month in culture, low molecular weight DNA was isolated from 1×10^7 cells and used to transform *Escherichia coli* DH5α, producing a total of 2475 hygromycin-resistant colonies.

Technical aspects of the library construction

An anti-ICP4 ribozyme library was produced as illustrated in Figure 3. To verify the effectiveness of this procedure 56 clones obtained at various steps were sequenced. Thirty-one were from the final ribozyme library and the remainder from earlier steps, beginning with the 14 bp fragment library. In some cases this led to modifications of the procedure to optimize the library construction. The results of the sequencing and the resulting procedural modifications are discussed below.

One observation made after the mung bean digestion was that the deletions infrequently stopped at A-T base pairs. While ExoIII has been shown to exhibit a preference for stopping at certain nucleotides (C > A = T > G) (45), this was not believed to be the cause of the observed sequence bias. Instead, we believe this was the result of a greater degree of 'breathing' at A-T terminated deletions and the subsequent removal of A-T terminal pairs by mung bean nuclease. The mung bean nuclease digestion was

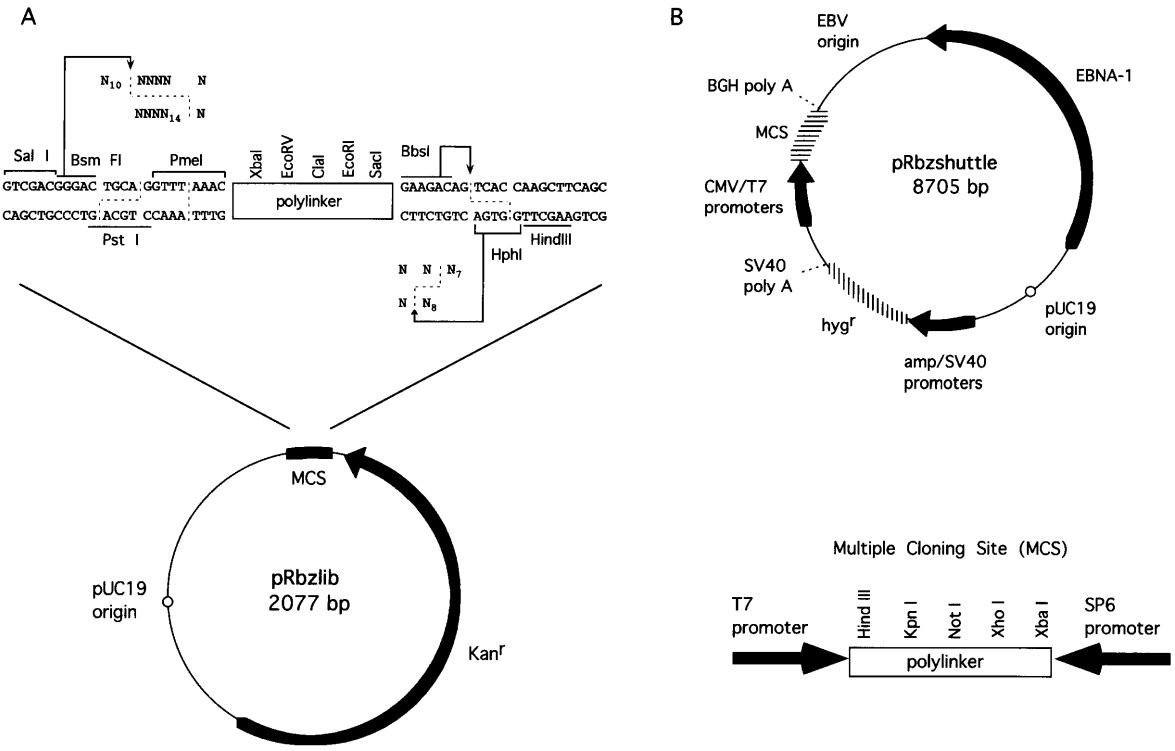

Figure 2. Plasmid vectors used in the construction and expression of the ribozyme library. (**A**) pRbzlib, used to construct the ribozyme library. The multiple cloning site is enlarged to show positions of important unique restriction endonuclease sites. (**B**) pRbzshuttle, used for expression of the ribozyme library in mammalian cells or by *in vitro* transcription using T7 RNA polymerase.

subsequently performed at higher salt concentrations (150 mM) and at a lower temperature (20°C). This eliminated the under-representation of A-T terminated deletions.

For construction of the library, two type IIS restriction enzymes are required, *Bsm*FI and *Hph*I. Typical of type IIS restriction enzymes, *Bsm*FI and *Hph*I cleave downstream of their recognition sequences in a sequence-independent manner (Fig. 2A). Cleavage by type IIS restriction enzymes can pose some problems since they can exhibit infidelity in how far from their recognition site they cleave. Cleavage by *Bsm*FI was largely at the expected distance (10/14), but also at 11/15. The 11/15 cleavage activity of the *Bsm*FI enzyme is the first published report of this type of infidelity. The reported 9/13 activity for this enzyme (46) was not seen in any of the clones sequenced. Infidelity by *Bsm*FI does not present a problem for construction of ribozyme libraries. The result of this infidelity is that the recognition domains of the ribozymes in the library can vary from 13 to 15 nt.

In contrast, *Hph*I infidelity can be problematical. *Hph*I digestion is a critical step in the construction of ribozyme libraries. This enzyme produces a 1 nt 3′-overhang which is subsequently removed by polishing with T4 DNA polymerase. It is essential to the proper functioning of the resulting ribozyme that this 1 nt is removed, since it does not have an antisense binding partner in the ribozyme (Fig. 1, H_{17}).

*Hph*I cleaves at 8/7, but also at 9/8 (47). This infidelity is demonstrated in our library by the presence of ribozymes with flanking helices of length 8 and 5, as would be expected if *Hph*I cleaved at 9/8. This type of infidelity, in itself, is not problematical. It simply alters the relative lengths of the two arms of the binding domain, leaving the total length of the binding arms unchanged. The problem that arises with *Hph*I infidelity is that the enzyme can cleave twice at the same target, i.e. if it first cleaves 9/8 it can rebind and cleave at 8/7. The result is that 2 bp are removed from the sequence upon subsequent polishing with T4 DNA polymerase. Removal of 2 bp from the insertion site of the ribozyme core cassette produces a non-functional ribozyme. In an early attempt to produce a library, >40% of the clones were the product of double cutting. This is close to the statistically predicted 50% that would result if *Hph*I has no preference for either 8/7 or 9/8 cutting. To minimize the possibility of double cutting, the *Hph*I digestion was performed under near 'single hit' conditions. Under these conditions double cleavage was reduced to 13% in the final library. It should be possible to further reduce the percentage of double hits by performing the cleavage under 'sub-single hit' conditions. This should not present any problems so long as the amount of plasmid digested is sufficient to allow full representation of the ribozyme library. Undigested molecules cannot accept the catalytic core and are removed in the subsequent step by selection for chloramphenicol resistance. Other class IIS restriction enzymes, such as *Mbo*II, could likely substitute for *Hph*I. However, it is not clear that their fidelity is any better.

The infidelity of *Hph*I raises another concern. It is possible that some sequences favor 8/7 and others 9/8 cutting. This could lead to the absence of some ribozyme target sequences in the final library. This appears to be unlikely. First, as discussed, under conditions that give nearly 100% cleavage by *Hph*I, >40% of the molecules are cut twice. This is close to the 50% predicted if *Hph*I exhibits no preference for 8/7 versus 9/8 cutting. Second, two clones that both contain the same 14 bp sequence of ICP4, Rz8 and Rz9

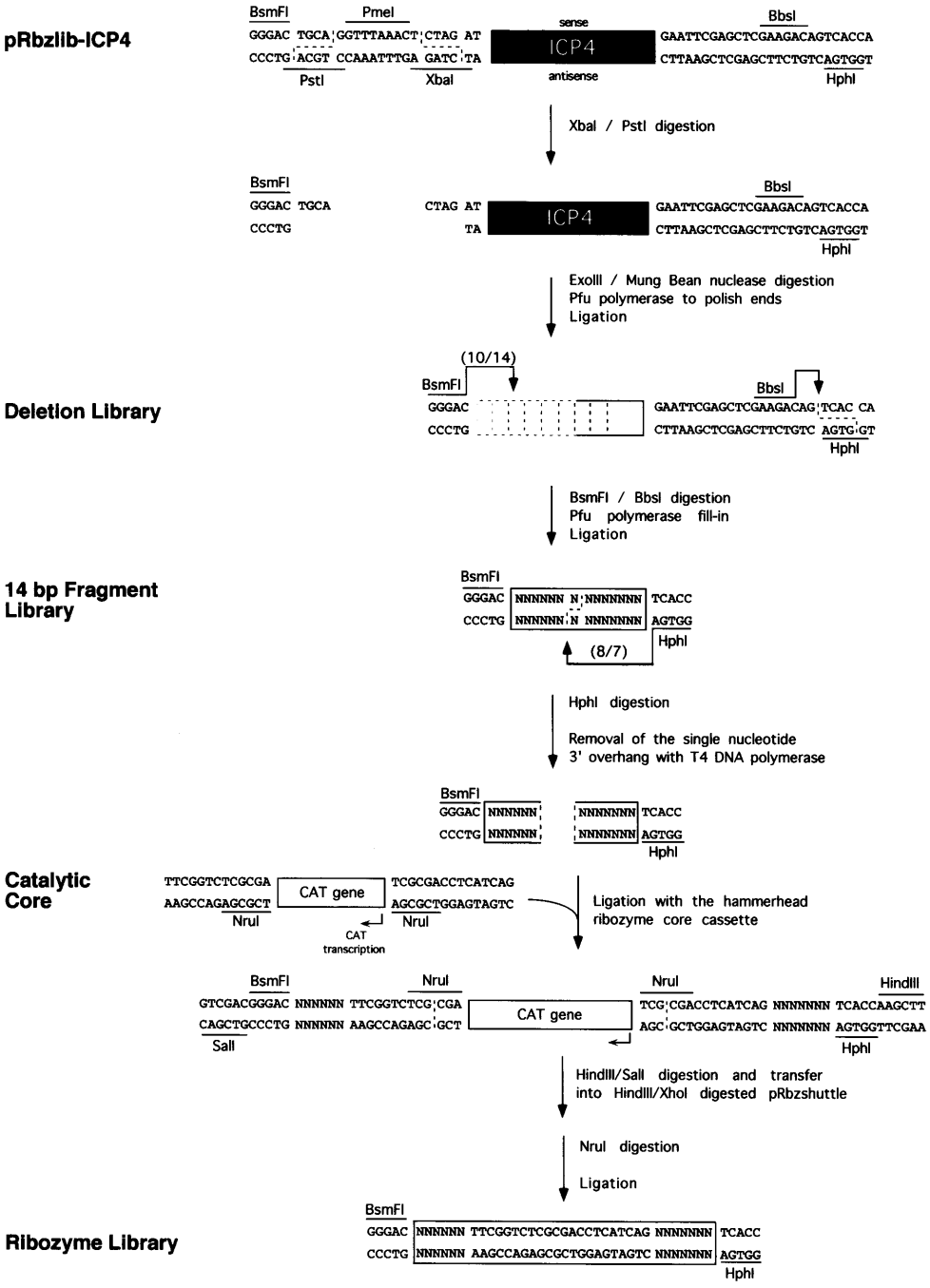

Figure 3. Construction of an anti-ICP4 ribozyme library. The library was constructed as illustrated and as described in Materials and Methods.

(Table 1), are the products of 8/7 and 9/8 cleavage, respectively. This suggests that the intervening sequence between the binding site and the cleavage site does not affect where *Hph*I cleaves.

*Hph*I is also sensitive to overlapping dam methylation. This is also true of *Mbo*II. Since 2 nt of the four base consensus sequence for dam methylation are provided by the variable sequence of the cDNA insert, mathematically 1/16 of the clones in the 14 bp fragment library (6.25%) will not be cleaved with *Hph*I and will be eliminated from the final ribozyme library. This can be prevented by passage of the 14 bp fragment library in a dam$^-$ strain prior to *Hph*I digestion.

Library distribution

The target locations of the 56 sequenced clones are illustrated in Figure 4. The histogram indicates that the target sites are fairly evenly distributed across the entire ICP4 gene, with the exception that no clones are identified targeting the very 5′- and 3′-termini.

It is unlikely that the library is devoid of members targeting these regions since the libraries are prepared with complexities far exceeding the total number of sites on the gene. It is even possible that target sites in these regions are similarly represented as those identified by the sequenced clones. Due to the small number of clones sequenced, it is likely that some larger gaps in the data could be observed even for a uniformly represented library, such as the gap between positions 966 and 1282. Sequencing additional clones would shed light on these possibilities.

Table 1. Target sites within ICP4 from the final library

Clone[a]	Target sequence[b]
Rz1 (1754)	CGACGcCGCCCGCC
Rz2 (1992)	CUGCGCgCGUGGC
Rz3 (2045)	GCGCCuGCGCGGGG
Rz4 (2252)	CGCCGCcGACGCGC
Rz5 (2411)	CCCCCuCCCCGCG
Rz6 (2517)	GUGGCcGUGUCGCG
Rz7 (2590)	GCCACaCGGCGGCG
Rz8 (2729)	CGCCGCgCGGUGCG
Rz9 (2729, repeated twice)	CGCCGcGCGGUGCG
Rz10 (2837)	CCCCCUgCGCGCCUC
Rz11 (2915)	GGUGGUgCUGUACUC
Rz12 (3246)	GGGCCCgCGGUGUC
Rz13 (3275, repeated three times)	CCUGGcGUGCGAGC
Rz14 (3569)	GGGGACCACCG**AC**gCCAUGGC
Rz15 (3680)	CGUGGCgCUGGGGC
Rz16 (3842)	CGGGAUuCGCUGGG[c]

[a]Nucleotide position number (in parentheses) indicates the location of ribozyme binding in the genomic fragment of ICP4.
[b]Underlining indicates the position that the required NUH sequence would need to be to produce an active ribozyme. Nucleotides in lower case indicate the unbound nucleotide, i.e. position 17 in Figure 1.
[c]Bona fide ribozyme target.

In the event that these regions are truly under-represented, this can be rectified by altering the conditions of the ExoIII digestion; specifically, more DNA can be removed during the early and late time points and less DNA removed during the middle time points. Alternatively, before combining the different time points after the mung bean digestion, the relative amounts of DNA in each could be determined by gel electrophoresis. Based on this, varying amounts of each time point could be combined to give a uniform distribution.

Of the 56 sequences determined, 42 (75%) occurred only once, while four occurred multiple times (Fig. 4). Three were only mildly over-represented, with two or three occurrences compared with the single occurrence for the majority of clones. The three positions were 2054 and 3246, with two occurrences each, and position 2729, with three occurrences. One position, 3275, was significantly over-represented, occurring seven times. Five of the occurrences were observed within the 32 clones sequenced from the final library and the other two were found at early stages of the construction. The over-representation of particular sites is likely caused by some local sequence and/or structure in the DNA

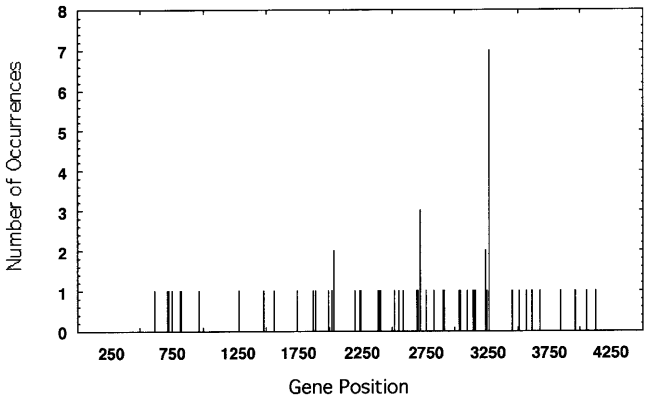

Figure 4. Distribution of 56 target sites identified in the ribozyme library. Nucleotide position numbering begins at the 5′ (+) strand end of the 4489 bp genomic fragment of ICP4 (the coding region runs from nt 126 to 4023). Repeated positions are 2045 (twice), 2729 (three times), 3246 (twice) and 3275 (seven times).

which either stalls ExoIII or causes it to fall off the template (48). Performing the ExoIII deletion at higher temperatures might reduce this if an inhibitory structure is forming at certain sequences. Higher temperature also allows for more distributive activity from the enzyme (49), which is desirable in this type of ExoIII digestion. While it is possible that the ExoIII digest conditions may need to be optimized for each target cDNA, creating libraries larger than would be necessary to represent every position would ensure complete representation of all target sites.

Library composition

Examination of the 31 clones obtained from the final library allowed determination of the overall effectiveness of the procedure. All 31 possessed a catalytic core demonstrating the effectiveness of the use of CAT selection for this purpose. Nineteen of the 31 clones (61%) contained sequences that could potentially be ribozymes, if the sequence that they target had included the required NUH sequence at the correct location. These are shown in Table 1. Counted amongst these potential ribozymes were three clones that possess non-detrimental defects. One has a single nucleotide deleted from loop II of the ribozyme (Rz13). This produces a 3, instead of 4 nt loop II. The site of this defect is the *Nru*I site used to remove CAT from the catalytic core. The ends must have been damaged during this step for this clone. The other two non-detrimental defects were the result of incomplete digestion by *Bsm*FI. These clones have a longer flanking arm corresponding to helix III (Rz12 and Rz14). This appears to be the result of a lack of cleavage of the *Bsm*FI site on pRbzlib and instead an internal *Bsm*FI site on ICP4 was used. These clones would be expected to produce functional ribozymes had they targeted an NUH sequence.

The remaining 12 clones (39% of 31) possessed defects that would prevent them from being potentially functional ribozymes. Four of these (13%) were defective in that they were cleaved twice with *Hph*I. As discussed above, it is likely that this defect can be reduced to close to zero, by performing the *Hph*I digestion under ‘sub-single-hit’ conditions. Three (9.7%) were missing 1 nt from one end of the catalytic core. Since the deletion always occurred at the same end of the cassette and the thermostable

polymerase used to make the cassette does not contain any 5′→3′ exonuclease activity, the PCR primer constituting that end of the cassette must have been contaminated with a small percentage of a failure fragment of the DNA synthesis. This defect can be eliminated by better purification of the primers. Five clones (16%) possessed the catalytic core in the incorrect orientation. This is in contrast to the expected 50% if there was no selection for orientation. With the current vector this number is not likely to be reduced. However, incorrectly oriented clones could be eliminated with a slight design change. If the promoter for the CAT gene were placed outside the MCS of pRbzlib, selection for correctly oriented catalytic cores should be 100%. Finally, three clones were the result of various unknown cloning artifacts. The number of defects described exceeds 12. This is due to the fact that, of the 12 defective clones, some had more than one of the defects.

Even with the 39% defect rate in our library, we are left with a success rate of 61%. As discussed, a few procedural changes should increase the success rate to 70–80%. This could be increased a further 16% by placing the CAT promoter outside the MCS. Even at 61%, this is still more than adequate. This just means that it is necessary to screen a ribozyme library 140% the size needed if 100% success were achieved. This would still be a small library relative to 'non-directed' library approaches.

Three out of the 31 clones (9.7%) targeted a site on the ICP4 mRNA that contained a uridine at the proper position of the consensus NUH site (Rz3, Rz5 and Rz16). Of the three, only one targeted a consensus NUH site (Rz16). Due to the unusually high G/C content of the ICP4 genomic fragment used to make the ribozyme library, only 9.2% of the nucleotides in the mRNA are uridines, of which 203 occur as an NUH triplet. The fact that the percentage of sequenced clones in the library targeting an NU site is virtually identical to the percentage of uridines in the ICP4 gene suggests that the library is unbiased and likely contains a fairly uniform distribution of target sites.

The use of a directed library for target site selection significantly simplifies the screening process, since only very small libraries need be prepared and assayed. For ICP4, assuming the library contains a uniform distribution of the 4475 distinct sequences (4489 – 14), a library of 67 125 (15-fold excess) is expected to have a probability of 99.9% of containing all sequences (50). Based on a χ^2 goodness-of-fit analysis of the 56 sequences, the multiples observed at positions 2729 and 3275 occur with a higher frequency than would be expected for a uniform distribution. All other positions are consistent with a uniform distribution. Correcting for the two over-represented sequences, a library of 81 057 (18-fold excess) is expected to contain all sequences with probability of 99.9%. Preparation, manipulation and screening of such a library is well within the limitations of current practice. In contrast, a non-directed library targeting 14 nt would require a minimum size of 2.7×10^8 (4^{14}). The ability to prepare and screen such a library is questionable. Even if possible, the vast majority of members of the library are directed at non-target genes. Inhibition of non-target genes could pose problems in interpreting the results.

Our expression vector, pRbzshuttle, was designed such that the identification of effective target sites could be performed *in vitro* or *in vivo*. For *in vitro* use, the ribozyme gene is flanked by a T7 promoter and an *Xba*I site (Fig. 2B). This allows preparation of the ribozyme library by *in vitro* transcription. Subsequently, the transcribed library can be used to challenge target mRNA, either prepared by *in vitro* transcription or present within a cell lysate. This approach has previously shown variable success in identifying effective target sites using a degenerate ribozyme library (22).

Although the *in vitro* use of degenerate libraries has shown some success, identification *in vivo* is expected to be of greater value. *In vivo* expression of an antisense library will ensure that the target mRNA is in its natural state and associated with its normal complement of factors. This will allow effective targets to be identified under the same conditions in which an antisense agent will need to function therapeutically. Additionally, some target sites may only be made accessible by the action of the many processing reactions that lead to maturation and expression of the mRNA. These sites could prove to be the more effective targets. These sites can likely only be identified in intact, fully functional cells. For the reasons discussed above, it is unlikely that a non-directed library can be assayed *in vivo*. In contrast, this should be easily obtainable with our directed ribozyme libraries due to their relatively small size and specific targeting of only the desired gene. This possibility is currently under examination.

While our libraries are based on the hammerhead ribozyme, sites identified should also be suitable for targeting by other antisense molecules, especially oligodeoxynucleotides that can activate RNase H-mediated cleavage. However, in a bacterial system, ribozyme libraries have identified effective sites that target non-NUH sequences (Z. Chen and D. E. Ruffner, unpublished observation). Although these sites are non-cleavable, gene inhibition requires the presence of the hammerhead catalytic core. Simple antisense RNAs targeted to the same sites are ineffective. Therefore, it appears that identification of effective targets is not limted to cleavable sites.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Robert Schackmann at the University of Utah oligonucleotide/peptide core facility for synthesizing all oligonucleotides used in this study. The core facility receives support from NCI grant CA42014. The authors also thank Dr Bill Sugden for helpful discussions regarding episomal shuttle vectors, Dr Saul Silverstein for his gift of plasmid pTEG2 and Louis Magas for statistical analysis. This work was funded by a grant from the NIH to D.E.R., 5R29AI34278.

REFERENCES

1 Cohen,J.S. (1991) *Pharmacol. Ther.*, **52**, 211–225.
2 Helene,C. and Toulme,J.J. (1990) *Biochim. Biophys. Acta*, **1049**, 99–125.
3 Englisch,U. and Gauss,D.H. (1991) *Angew. Chem.*, **30**, 613–722.
4 Uhlmann,E. and Peyman,A. (1990) *Chem. Rev.*, **90**, 543–584.
5 James,W. and Cowe,E. (1997) *Methods Mol. Biol.*, **74**, 17–26.
6 Jarvis,T.C., Wincott,F.E., Alby,L.J., McSwiggen,J.A., Beigelman,L., Gustofson,J., DiRenzo,A., Levy,K., Arthur,M., Matulic Adamic,J., Karpeisky,A., Gonzalez,C., Woolf,T.M., Usman,N. and Stinchcomb,D.T. (1996) *J. Biol. Chem.*, **271**, 29107–29112.
7 Sczakiel,G. and Tabler,M. (1997) *Methods Mol. Biol.*, **74**, 11–15.
8 Xing,Z. and Whitton,J.L. (1992) *J. Virol.*, **66**, 1361–1369.
9 Lima,W.F., Monia,B.P., Ecker,D.J. and Freier,S.M. (1992) *Biochemistry*, **31**, 12055–12061.
10 Dreyfuss,G. (1986) *Annu. Rev. Cell. Biol.*, **2**, 459–498.
11 Perriman,R., Delves,A. and Gerlach,W.L. (1992) *Gene*, **113**, 157–163.
12 Zoumadakis,M. and Tabler,M. (1995) *Nucleic Acids Res.*, **23**, 1192–1196.
13 Clouet d'Orval,B. and Uhlenbeck,O.C. (1997) *Biochemistry*, **36**, 9087–9092.
14 Fedor,M.J. and Uhlenbeck,O.C. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 1668–1672.
15 Bertrand,E.L. and Rossi,J.J. (1994) *EMBO J.*, **13**, 2904–2912.

16 Clouet d'Orval,B., Stage,T.K. and Uhlenbeck,O.C. (1995) *Biochemistry*, **34**, 11186–11190.
17 Herschlag,D., Khosla,M., Tsuchihashi,Z. and Karpel,R.L. (1994) *EMBO J.*, **13**, 2913–2924.
18 Sioud,M. and Jespersen,L. (1996) *J. Mol. Biol.*, **257**, 775–789.
19 Stage,T.K., Hertel,K.J. and Uhlenbeck,O.C. (1995) *RNA*, **1**, 95–101.
20 Tsuchihashi,Z., Khosla,M. and Herschlag,D. (1993) *Science*, **262**, 99–102.
21 Koguma,T., Ohkawa,J., Mori,K., Fujita,S., Nishikawa,S., Iwakura,M. and Taira,K. (1995) *Nucleic Acids Res. Symp. Ser.*, **34**, 221–222.
22 Lieber,A. and Strauss,M. (1995) *Mol. Cell. Biol.*, **15**, 540–551.
23 Campbell,T.B. and Cech,T.R. (1995) *RNA*, **1**, 598–609.
24 Frank,B.L. and Goodchild,J. (1997) *Methods Mol. Biol.*, **74**, 37–43.
25 Birikh,K.R., Berlin,Y.A., Soreq,H. and Eckstein,F. (1997) *RNA*, **3**, 429–437.
26 Branch,A.D. (1998) *Antisense Nucleic Acid Drug Dev.*, **8**, 249–254.
27 Matveeva,O., Felden,B., Audlin,S., Gesteland,R.F. and Atkins,J.F. (1997) *Nucleic Acids Res.*, **25**, 5010–5016.
28 Ruffner,D.E., Stormo,G.D. and Uhlenbeck,O.C. (1990) *Biochemistry*, **29**, 10695–10702.
29 Shimayama,T., Nishikawa,S. and Taira,K. (1995) *Biochemistry*, **34**, 3649–3654.
30 International Union of Biochemistry. (1985) *Eur. J. Biochem.*, **150**, 1–5.
31 Zhu,X.X., Papavassiliou,A.G., Stunnenburg,H.G. and Silverstein,S. (1991) *Virology*, **184**, 67–78.
32 Allen,L.N. and Hanson,R.S. (1985) *J. Bacteriol.*, **161**, 955–962.
33 Merino,E., Osuna,J., Bolivar,F. and Soberon,X. (1992) *Biotechniques*, **12**, 508–510.
34 Horinouchi,S. and Weisblum,B. (1982) *J. Bacteriol.*, **150**, 815–825.
35 Damak,S. and Bullock,D.W. (1993) *Biotechniques*, **15**, 448–450.
36 Dixon,R.F. and Schaffer,P.A. (1980) *J. Virol.*, **36**, 189–203.
37 DeLuca,N.A., McCarthy,A.M. and Schaffer,P.A. (1985) *J. Virol.*, **56**, 558–570.
38 DeLuca,N.A. and Schaffer,P.A. (1988) *J. Virol.*, **62**, 732–743.
39 Bruckner,R., Zyprian,E. and Matzura,H. (1984) *Gene*, **32**, 151–160.
40 Yates,J., Warren,N., Reisman,D. and Sugden,B. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 3806–3810.
41 Yates,J.L., Warren,N. and Sugden,B. (1985) *Nature*, **313**, 812–815.
42 Belt,P.B., Groeneveld,H., Teubel,W.J., van de Putte,P. and Backendorf,C. (1989) *Gene*, **84**, 407–417.
43 Yates,J.L. and Guan,N. (1991) *J. Virol.*, **65**, 483–488.
44 Margolskee,R.F., Kavathas,P. and Berg,P. (1988) *Mol. Cell. Biol.*, **8**, 2837–2847.
45 Linxweiler,W. and Horz,W. (1982) *Nucleic Acids Res.*, **10**, 4845–4859.
46 Velculescu,V.E., Zhang,L., Vogelstein,B. and Kinzler,K.W. (1995) *Science*, **270**, 484–487.
47 Kleid,D., Humayun,Z., Jeffrey,A. and Ptashne,M. (1976) *Proc. Natl Acad. Sci. USA*, **73**, 293–297.
48 Abarzua,P. and Marians,K.J. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 2030–2034.
49 Hoheisel,J.D. (1993) *Anal. Biochem.*, **209**, 238–246.
50 Feller,W. (1968) *An Introduction to Probability Theory and Its Applications*, 3rd Edn. John Wiley & Sons, New York, NY.