OLIGONUCLEOTIDES 14:210–220 (2004)

ROLL: A Method of Preparation of Gene-Specific Oligonucleotide Libraries

ALEXANDER V. VLASSOV,[1] OLGA A. KOVAL,[2] BRIAN H. JOHNSTON,[1] and SERGEI A. KAZAKOV[1]

ABSTRACT

The selection of nucleic acid sequences capable of specifically and efficiently hybridizing to target sequences is crucial to the success of many applications, including microarrays, PCR and other amplification procedures, antisense inhibition, ribozyme-mediated cleavage, and RNA interference (RNAi). Methods of selection using nucleotide sequence libraries have several advantages over rational approaches using defined sequences. However, the high complexity of completely random (degenerate) libraries and their high toxicity in cell-based assays make their use in many applications impractical. Gene-specific oligonucleotide libraries, which contain all possible sequences of a certain length occurring within a given gene, have much lower complexity and, thus, can significantly simplify and accelerate sequence screening. Here, we describe a new method for the preparation of gene-specific libraries using the ligation of randomized oligonucleotide probes hybridized adjacently on target polynucleotide templates followed by PCR amplification. We call this method random oligonucleotide ligated libraries (ROLL).

INTRODUCTION

THE MOST IMPORTANT REQUIREMENTS for diagnostic and therapeutic technologies based on hybridization between probe and target nucleic acids are high sequence specificity and strong probe-target interactions. Such technologies include microarrays (Southern et al., 1999), competitive RT-PCR (Ishibashi, 1997), and ligation-mediated amplification (Landegren et al., 1996) assays, as well as antisense (Bruice and Lima, 1997; Sohail and Southern, 2000), ribozyme (Amarzguioui et al., 2000; Scarabino and Tocchini-Valentini, 1996), and small interfering RNA (siRNA) (Holen et al., 2002) approaches to gene inhibition. The specificity and efficacy of probe-target interactions depend on such parameters as target accessibility, hybridization rate, and duplex stability (Sczakiel and Far, 2002). Because of the complexity of these interactions, rational design methods for predicting optimal probe sequences and target site accessibility have had limited success (Sczakiel and Far, 2002; Sohail and Southern, 2000). Also, the common notion that sequences that are less involved in internal hydrogen bonding interactions represent more favorable target sequences is a clear oversimplification (Fakler et al., 1994; Laptev et al., 1994; Sczakiel and Far, 2002). Target RNAs are often folded differently in the cell than *in vitro* (Lindell et al., 2002) and may be complexed with proteins that reduce target site accessibility (Lieber and Strauss, 1995). Conversely, some cellular factors may promote probe hybridization with target sites that are not accessible *in vitro* (Bertrand and Rossi, 1994; Laptev et al., 1994).

[1]SomaGenics, Inc., Santa Cruz, CA 95060.
[2]Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk 630090, Russia.

As a consequence of this complexity, optimal hybridizing sequences cannot be rationally selected based on either sequence data or predicted or experimentally determined target accessibility. To address this problem, several *in vitro* and *in vivo* methods for selecting target sequences from sequence libraries with 5–30 variable nucleotides have been developed (Allawi et al., 2001; Birikh et al., 1997; Ho et al., 1998; Liang et al., 2002; Lieber and Strauss, 1995; Lima et al., 1997; Lloyd et al., 2001; Milner et al., 1997; Paquin et al., 2000; Patzel and Sczakiel, 2000; Scherr et al., 2001; Wrzesinski et al., 2000; Yu et al., 1998). An additional advantage of such libraries is that they can be used in a "reverse genomics" approach, which can identify genes responsible for a specific phenotype without prior knowledge of any sequence information (Kawasaki and Taira, 2002; Li et al., 2000).

In principle, screening for gene inhibitors as well as diagnostic and affinity probes may be performed by using completely random (degenerate) libraries (Kruger et al., 2000; Lieber and Strauss, 1995). This approach has several major problems, however. First, the high complexity of random libraries does not allow them to be screened in mammalian cells with a single sequence represented in each cell (Yu et al., 1998). Indeed, for 18-nt antisense sequences, the library complexity is 4^{18} or $\sim 10^{11}$ molecules, which requires that 10^{11} cells should be screened. Second, as each sequence in a degenerate library has its complement also represented, the two can form stable duplexes, thus reducing their availability for interaction with accessible target sites (Ho et al., 1996). Third, experiments have shown that degenerate libraries are highly toxic to cells; antisense ribozymes with degenerate substrate recognition sites can efficiently block the functioning of both mRNAs of interest (host or foreign) and unintended cellular RNAs (Kruger et al., 2000; Pierce and Ruffner, 1998). Gene-specific, directed libraries (Pierce and Ruffner, 1998), comprising only sequences represented within the target gene(s) of interest, offer a superior alternative to screening completely random libraries. The use of directed libraries, prepared or preselected *in vitro*, significantly simplifies the screening process, as comparatively small libraries need to be assayed. For example, an 18-nt directed library targeting an mRNA of 2000 nt consists of only 1983 different molecules. The reduced complexity of the directed libraries is also beneficial for producing good hybridization within a reasonable time (Southern et al., 1999). Further, unintended knockdown of nontargeted genes would be reduced, allowing more efficient cell-based assays with the directed libraries cloned into appropriate vectors.

Currently, several methods of preparation of directed libraries (which could be amplified and cloned) are available. The first method published, developed by Pierce and co-workers (Pierce and Ruffner, 1998; Ruffner et al., 1999) specifically for the hammerhead ribozyme constructs, involves multiple enzymatic manipulations to produce a directed library of antisense sequences with a uniform length (14 nt). In addition to the technical complexity of the procedure, this method has additional disadvantages in that the terminal ~500 nt at each end of the target sequences are missing, and the size of the antisense sequences is restricted to a 14-nt length (which is presumably less than required for specific targeting of human genes) because of the limitations of the enzymatic procedures used. Another method, developed by Paquin's group (Paquin et al., 2000; Brukner et al., 2002) is based on simple screening of completely random 20-nt sequences to select only those that hybridize to the immobilized target DNA. These randomized sequences are placed between two defined primer-binding sequences that are used for PCR amplification of the selected hybridizing sequences. This method is comparatively simple but suffers from a high level of nonspecific hybridization, resulting in five or six mismatches in the majority of the 20-mer sequences selected. In fact, the use of random hybridization probes longer than 15–16 nt was found to be counterproductive (Lloyd et al., 2001) because self-pairing interactions limit the number of nucleotides available for hybridization.

Yet another method based on template-assisted combinatorial strategy was described by Boiziau et al. (1999), who selected DNA aptamers targeting an accessible binding site in an RNA hairpin using both completely random libraries and libraries enriched in target-specific sequences. The enriched sequences were produced by ligation of half-candidates in the presence of an RNA hairpin using RNA ligase. The half-candidates were designed as hemirandom probes containing defined primer and comparatively long, 15-nt terminal random sequences. The ligation method showed low efficiency and target specificity, which is a consequence of the preference of RNA ligase to ligate sequence motifs that are not aligned in complementary complexes (Harada and Orgel, 1993). Also, most ligation products were unrelated to the RNA target, and the authors found no benefit to using libraries prepared from hemirandom probes vs. using probes with completely random 30-mer libraries without a ligation step.

Recently, Shirane et al. (2004) developed another method for preparing directed libraries of 19–21 bp DNA fragments and their use in vectors to express short interfering RNA libraries. This method includes fragmentation of the DNA of interest by DNase I to an average size of 100–200 bp, blunt-ending with DNA polymerase, and ligating to a hairpin-shaped adaptor containing the recognition sequence of *Mme*I restriction endonuclease. Subsequent cleavage by *Mme*I produced DNA fragments of 19–21 bp. This preparation scheme is rather complex,

and the library obtained is restricted in length to ~20 nt. In a similar approach, Sen et al. (2004) used a mixture of restriction endonucleases producing CG overhangs for fragmenting the target DNA. This cocktail does not produce sufficiently random cuts, and as a result, the obtained library contained only 34 unique target-specific sequences out of the theoretically possible 981 for the 1000-nt long target.

Thus, there is a need for an improved method for generating gene-specific oligonucleotide libraries that does not suffer from the limitations of the methods described. The approach described here involves ligation of oligonucleotides containing randomized segments that are hybridized to adjacent sites on the target polynucleotide templates. The method, random oligonucleotide ligated libraries (ROLL), requires two oligonucleotides consisting of fixed primer/restriction-site sequences linked to randomized 10-nt sequences at their 5′-end or 3′-end, respectively (Fig. 1A). When a pair of these oligonucleotides happens to be complementary to adjacent sites on a target DNA or RNA, they can be ligated by the addition of DNA ligase, thereby linking the two fixed sequences and allowing PCR amplification. Similar events happening at all accessible sequences on the target generate a directed library specific to the chosen target. Both the hybridization and the ligation of these short randomized sequences provide higher sensitivity to mismatches than methods based on hybridization alone of longer sequences. After PCR amplification of the ligated probes, the primer-binding sequences can be cleaved off by restriction enzymes, and the library can be inserted into expression cassettes. From these cassettes, the gene-specific libraries of antisense, ribozyme, or siRNAs can be expressed, and using various selection procedures, the best binders or inhibitors to a particular target can be screened and identified.

MATERIALS AND METHODS

Synthetic oligonucleotide probes and primes

All oligodeoxyribonucleotides used in this work were synthesized by Integrated DNA Technologies (Coralville, IA). Oligonucleotides were gel purified (15% PAGE/7 M urea), extracted from the gel, precipitated with ethanol, and then dissolved in deionized water. The sequences (5′–3′) of hemirandom probes were pNNNNNNNNNNGGATCCCTGCTGACGACTAGACTGTG (I) and CAGTCTAGCAAGTATGCGTCCTCGAGNNNNNNNNNN (II). Masking oligonucleotides complementary to the predetermined sequences of the hemirandom probes were CACAGTCTAGTCGTCAGCAGGGATCC (I) and CTCGAGGACGCATACTTGCTAGACTG (II), and PCR primers were CACAGTCTAGTCGTCAGCAG (primer 1) and CAG-TCTAGCAAGTATGCGTC (primer 2). Random oligoribonucleotides 4, 5, 6, and 7 nt in length (used in competition experiments) were obtained from Dharmacon (Lafayette, CO).

DNA targets (ligation templates)

The Semliki Forest virus (SFV) target consisted of two DNA molecules, the SFV replicon DNA (7378 bp) and the SFV helper DNA (5092 bp) (sequences available in Garoff et al., 1980; Takkinen, 1986). These DNAs were obtained by PCR from plasmids (kindly provided by Dr. Kenneth Lundstrom) heat denatured (95°C for 2 minutes, then quenched in ice) prior to use as ligation templates. The mouse tumor necrosis factor (TNF-α) DNA was obtained by PCR from plasmid provided by Belgian Coordinated Collections of Microorganisms (sequences available in Fransen et al., 1985; Pennica et al., 1985). The single-stranded DNA (ssDNA) was further obtained by asymmetric PCR (10× excess of one primer over another). In early experiments, we also used two different 1-kb fragments of the SFV genome DNA (heat denatured) and several short (20–30 nt) single-stranded oligonucleotide targets).

Hybridization and ligation of hemirandom probes on the DNA target

The hemirandom probes I and II were prehybridized with their corresponding masking oligonucleotides at 25°C for 5 minutes in 30 mM Tris-HCl, pH 7.8, buffer containing 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 50 mM NaCl. Then, the denatured target DNA was added, and the mixture was further incubated for 30 minutes to allow the probes to bind the DNA templates. Finally, T4 DNA ligase (Promega, Madison, WI) was added, and the mixture was incubated at 25°C for 1 hour.

Typical reaction conditions were as follows: hemirandom probes I and II, 0.1–1 μM (2–20 pmol); masking oligonucleotides, 0.1–1 μM (2–20 pmol); DNA target, 0.01–1 μM (0.2–20 pmol); T4 DNA ligase, 0.1 U/μl (2 U). The final volume was 20 μl.

In control experiments, we performed the ligation reactions omitting the masking oligonucleotides, omitting DNA target, or using hemirandom probes with different or shorter predetermined sequence or shorter random sequence, as well as at different temperatures (from 20°C to 40°C) and salt concentrations (2–10 mM $MgCl_2$ and 50–200 mM NaCl).

PCR amplification of ligated products

Each ligation mixture (1 μl) was used for PCR amplification of the ligation products in solutions containing 5 μl

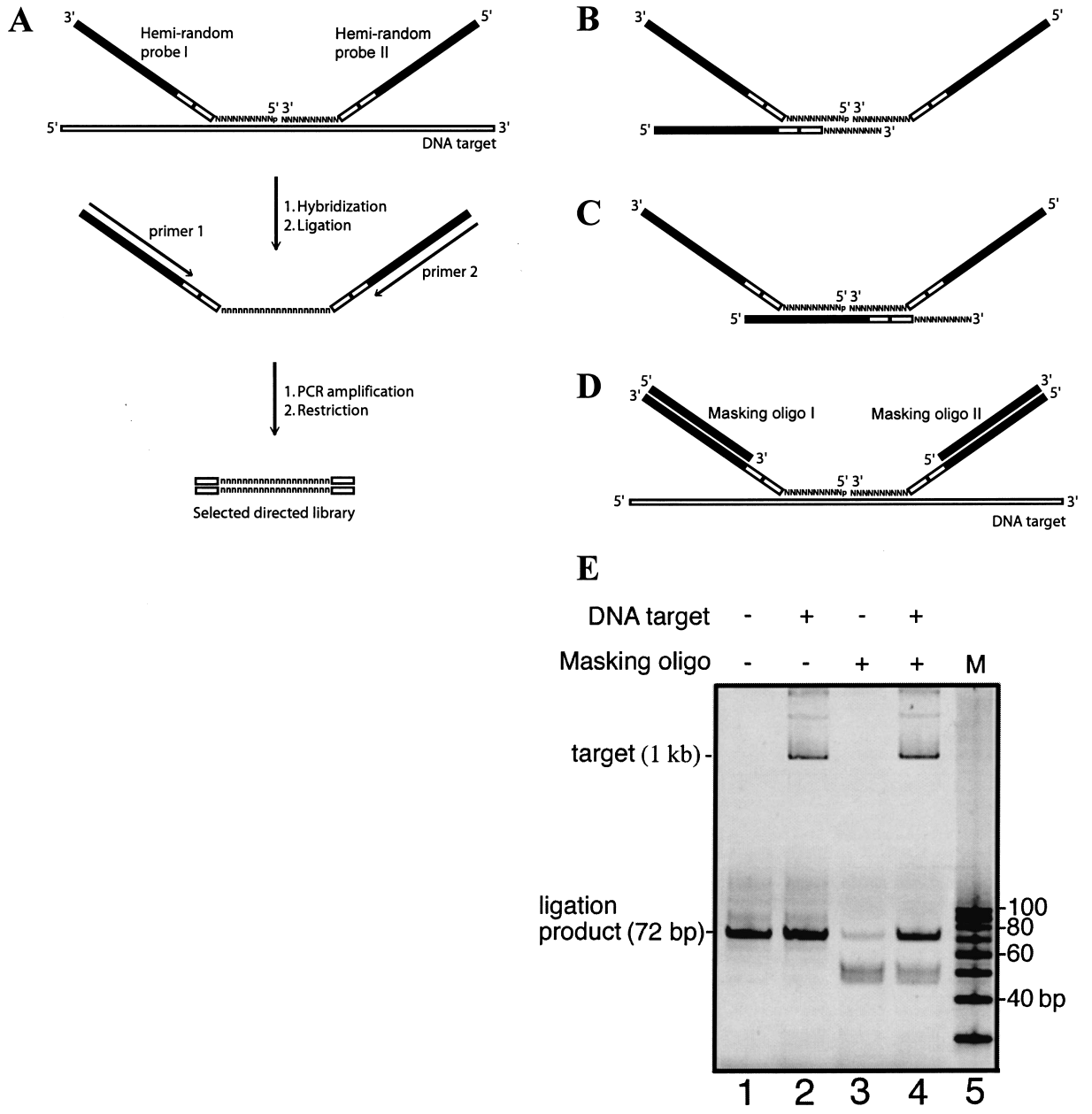

FIG. 1. Preparation of gene-specific libraries by ligating two hemirandom probes hybridized to an ssDNA target. (**A**) General experimental scheme. (**B** and **C**) Possible structures leading to target-independent self-ligation. (**D**) Masking oligonucleotides that inhibit target-independent ligation. (**E**) Electrophoretic analysis of the PCR-amplified ligation products on a nondenaturing 10% polyacrylamide gel stained with ethidium bromide (negative image). Both unmasked (lanes 1 and 2) and masked (lanes 3 and 4) hemirandom probes I and II were ligated in the absence (lanes 1 and 3) or presence (lanes 2 and 4) of a 1-kb DNA target. M, molecular weight markers.

10× PCR buffer, 3 μl of 25 mM $MgCl_2$, 4 μl of dNTPs (2.5 mM), 0.5 μl primer 1 (50 μM), 0.5 μl primer 2 (50 μM), 0.25 μl Taq DNA polymerase (5 U/μl) (Promega), and sufficient water to bring the volume to 50 μl. The typical PCR cycle was 94°C/30 seconds, 54°C/30 seconds, 72°C/15 seconds, repeated for 15–20 cycles.

Gel analysis of PCR products

PCR reaction (5–10 μl) was mixed with 2 μl of 6× loading buffer (0.25% bromphenol blue, 0.25% xylene cyanol, 30% glycerol in water) and electrophoresed through a nondenaturing 10% polyacrylamide gel at 25

A

1 **CAGTCTAGCAAGTATGCGTCCTCGAG**TCGCGGTCCT-AGATTGGTGT**GGATCCCTGCTGACGACTAGACTGTG**

2 **CAGTCTAGCAAGTATGCGTCCTCGAG**GGGGGTTGAG-TGGCTGGTCT**GGATCCCTGCTGACGACTAGACTGTG**

3 **CAGTCTAGCAAGTATGCGTCCTCGAG**GAGGCAGGGT-CCTACTAAAT**GGATCCCTGCTGACGACTAGACTGTG**

4 **CAGTCTAGCAAGTATGCGTCCTCGAG**GGTTCTAGAT-TGGTGCTTCT**GGATCCCTGCTGACGACTAGACTGTG**

5 **CAGTCTAGCAAGTATGCGTCCTCGAG**AAGAAGCGGT-CCTAGATTGA**GGATCCCTGCTGACGACTAGACTGTG**

6 **CAGTCTAGCAAGTATGCGTCCTCGAG**GTGTGGTGCA-AGGGATATAC**GGATCCCTGCTGACGACTAGACTGTG**

7 **CAGTCTAGCAAGTATGCGTCCTCGAG**AAGAAGGGGG-GGATGGTTGC**GGATCCCTGCTGACGACTAGACTGTG**

8 **CAGTCTAGCAAGTATGCGTCCTCGAG**GCGGTAACTA-GATTGGTGTC**GGATCCCTGCTGACGACTAGACTGTG**

9 **CAGTCTAGCAAGTATGCGTCCTCGAG**TAATCGTATC-ATGGGAGGGA**GGATCCCTGCTGACGACTAGACTGTG**

10 **CAGTCTAGCAAGTATGCGTCCTCGAG**TCGTTGGGTG-AAGGGCTAGA**GGATCCCTGCTGACGACTTGACTGTG**

11 **CAGTCTAGCAAGTATGCGTCCTCGAG**GATCGGGAAG-GGTACGACCG**GGATCCCTGCTGACGACTAGACTGTG**

12 **CAGTCTAGCAAGTATGCGTCCTCGAG**TGGAGGCCTA-GAGTGGATAC**GGATCCCTGCTGACGACTTGACTGTG**

13 **CAGTCTAGCAAGTATGCGTCCTCGAG**GCACTGGGTA-GTCGTGAACC**GGATCCCTGCTGACGACTAGACTGTG**

14 **CAGTCTAGCAAGTATGCGTCCTCGAG**AAGTGCTGCA-TACGATATCC**GGATCCCTGCTGACGACTAGACTGTG**

15 **CAGTCTAGCAAGTATGCGTCCTCGAG**GAGATGGGCT-CCATGGGCTT**GGATCCCTGCTGACGACTAGACTGTG**

16 **CAGTCTAGCAAGTATGCGTCCTCGAG**CATACGCCGC-ATAGTGTTCA**GGATCCCTGCTGACGACTAGACTGTG**

17 **CAGTCTAGCAAGTATGCGTCCTCGAG**TCACTGCTGC-GACAAATGAA**GGATCCCTGCTGACGACTAGACTGTG**

B

1 **CAGTCTAGCAAGTATGCGTCCTCGAG**ATCACACGGG-TGTACTACAG**GGATCCCTGCTGACGACTAGACTGTG**

2 **CAGTCTAGCAAGTATGCGTCCTCGAG**CGGTCTAGTC-GACATATTGG**GGATCCCTGCTGACGACTAGACTGTG**

3 **CAGTCTAGCAAGTATGCGTCCTCGAG**GAGGAATGGT-GGCTCACGAA**GGATCCCTGCTGACGACTAGACTGTG**

4 **CAGTCTAGCAAGTATGCGTCCTCGAG**TATGTTTGAT-TAAAAAATGA**GGATCCCTGCTGACGACTAGACTGTG**

5 **CAGTCTAGCAAGTATGCGTCCTCGAG**TAAGAGTCTA-GTGTGCGGCC**GGATCCCTGCTGACGACTAGACTGTG**

FIG. 2. Representative sequences from SFV-specific libraries generated under various ligation conditions. The 3′- and 5′-flanking sequences (26 nt, shown in bold) are the primer-binding sites used for PCR amplification of the ligated probes. The central 20-nt segment was derived from the 10-nt randomized regions of hemirandom probes I and II after ligation. The ligation site is indicated by a dash (–). The target-matching sequence is underlined. (**A**) Ligation performed in the presence of target, with masking oligos, at 25°C, and buffer containing 10 mM Mg^{2+} and 50 mM Na^{+}. (**B**) Ligation performed in the absence of target, without masking oligos, at 25°C, and buffer containing 10 mM Mg^{2+} and 50 mM Na^{+}. (*Continued on the next page*)

V/cm field. The gel was stained with ethidium bromide and analyzed using a PhosphorImager (Molecular Imager FX, Bio-Rad, Hercules, CA).

Cloning and sequencing

The PCR-amplified 72-bp products were gel purified and ligated into the pT7Blue-3 vector (Novagen, Madison, Wisconsin). *Escherichia coli* competent cells (Novagen) were transformed with the recombinant vector, and colonies were selected for minipreps. DNA was isolated using the QIAprep Spin Miniprep Kit (Qiagen, Chatsworth, CA) and sequenced at the Marshall University DNA Core Facility (Huntington, WV).

RESULTS AND DISCUSSION

Design of hemirandom probes and scheme of directed library production

For preparation of the gene-specific oligonucleotide library, two hemirandom probes were designed (Fig. 1A). Both hemirandom probes contained predetermined sequences consisting of primer-binding sequences with restriction sites plus a 10-nt randomized sequence located either at the 5′-end (probe I) or 3′-end (probe II). To satisfy the substrate requirements for DNA ligase, the 5′-end of probe I was phosphorylated, whereas probe II had a 3′-OH end. The hemirandom probes were hybridized

C

1 **CAGTCTAGCAAGTATGCGTCCTCGAG**CGTTTTCGGC-TTTGGGTTGG**GGATCCCTGCTGACGACTAGACTGTG**

2 **CAGTCTAGCAAGTATGCGTCCTCGAG**GCGGTCCTAG-ATGGGTGCGC**GGATCCCTGCTGACGACTAGACTGTG**

3 **CAGTCTAGCAAGTATGCGTCCTCGAG**TATGGTATGG-GTTTGTTGGT**GGATCCCTGCTGACGACTAGACTGTG**

4 **CAGTCTAGCAAGTATGCGTCCTCGAG**CTTTTGTCGA-ATGATGGCGG**GGATCCCTGCTGACGACTAGACTGTG**

5 **CAGTCTAGCAAGTATGCGTCCTCGAG**AGCGCTTCCG-AGGCTGGGGG**GGATCCCTGCTGACGACTAGACTGTG**

6 **CAGTCTAGCAAGTATGCGTCCTCGAG**CGGCAGGTGG-CTGGAGCTGT**GGATCCCTGCTGACGACTAGACTGTG**

7 **CAGTCTAGCAAGTATGCGTCCTCGAG**ATGGGCCTAG-ATTGGGGCGC**GGATCCCTGCTGACGACTAGACTGTG**

8 **CAGTCTAGCAAGTATGCGTCCTCGAG**CTAGCAAGTA-TGCGTGCTCG**GGATCCCTGCTGACGACTAGACTGTG**

9 **CAGTCTAGCAAGTATGCGTCCTCGAG**CCGCGATCAT-GGGGAGGGTC**GGATCCCTGCTGACGACTAGACTGTG**

10 **CAGTCTAGCAAGTATGCGTCCTCGAG**CGCGGTCCTA-AATTGTTGCC**GGATCCCTGCTGACGACTAGACTGTG**

D

1 **CAGTCTAGCAAGTATGCGTCCTCGAG**CCCGTTGGTA-GGTGAGAGCT**GGAT**CCCTGCTGACGACTAGACTGTG

2 **CAGTCTAGCAAGTATGCGTCCTCGAG**GGGATTGTAG-GTGCAGACAG**GGATCCCTGCTGACGACTAGACTGTG**

3 **CAGTCTAGCAAGTATGCGTCCTCGAG**TTGGCTGCAG-GGAGCGAATC**GGATCCCTGCTGACGACTAGACTGTG**

4 **CAGTCTAGCAAGTATGCGTCCTCGAG**CGCGTCCTGG-ATTGGTGTCC**GGATCCCTGCTGACGACTAGACTGTG**

5 **CAGTCTAGCAAGTATGCGTCCTCGAG**CGACAGTCTA-GTTGGCTTTT**GGATCCCTGCTGACGACTAGACTGTG**

E

1 **CAGTCTAGCAAGTATGCGTCCTCGAG**CCAGCGCGTG-CCAGTTGGTA**GGATCCCTGCTGACGACTAGACTGTG**

2 **CAGTCTAGCAAGTATGCGTCCTCGAG**CAAGTGGTTC-CGGACTCGTG**GGATCCCTGCTGACGACTAGACTGTG**

3 **CAGTCTAGCAAGTATGCGTCCTCGAG**GCGGTCCTAG-ATTGGGCCAG**GGATCCCTGCTGACGACTAGACTGTG**

4 **CAGTCTAGCAAGTATGCGTCCTCGAG**GGGTGTCCTG-GCTTTTGTCG**GGATCCCTGCTGACGACTAGACTGTG**

5 **CAGTCTAGCAAGTATGCGTCCTCGAG**TTGGGGATCC-CGGTAGGTTC**GGATCCCTGCTGACGACTAGACTGTG**

6 **CAGTCTAGCAAGTATGCGTCCTCGAG**GTTGGGATCC-AATTATTCCC**GGATCCCTGCTGACGACTAGACTGTG**

7 **CAGTCTAGCAAGTATGCGTCCTCGAG**CACAACGCTA-GTCGTTGGAC**GGATCCCTGCTGACGACTAGACTGTG**

8 **CAGTCTAGCAAGTATGCGTCCTCGAG**AGGAGGGGTC-CGGTCTTTAA**GGATCCCTGCTGACGACTAGACTGTG**

9 **CAGTCTAGCAAGTATGCGTCCTCGAG**GGCTGGGATC-CATGCTCTCT**GGATCCCTGCTGACGACTAGACTGTG**

F

1 **CAGTCTAGCAAGTATGCGTCCTCGAG**AAGCAGAGGT-GGCAGGAGCT**GGAT**CCCTGCTGACGACTAGACTGTG

2 **CAGTCTAGCAAGTATGCGTCCTCGAG**AGTCTCTCGC-GTAGCGGAGG**GGATCCCTGCTGACGACTAGACTGTG**

3 **CAGTCTAGCAAGTATGCGTCCTCGAG**GGTCTTTTGG-CGTCGTGTAC**GGATCCCTGCTGACGACTAGACTGTG**

4 **CAGTCTAGCAAGTATGCGTCCTCGAG**TTGTCTGGGT-CGCAAGGTTT**GGATCCCTGCTGACGACTAGACTGTG**

5 **CAGTCTAGCAAGTATGCGTCCTCGAG**ACTCAAGTAT-GCGTTCTTGA**GGATCCCTGCTGACGACTAGACTGTG**

6 **CAGTCTAGCAAGTATGCGTCCTCGAG**TCTAGCAGCA-TGCGTCCTCG**GGAGCCCTGCTGACGACTAGACTGTG**

FIG. 2. (*Continued from previous page*) (**C**) Ligation performed in the presence of target, with masking oligos, at 35°C, and buffer containing 10 mM Mg^{2+} and 50 mM Na^{+}. (**D**) Ligation performed in the presence of target, without masking oligos, at 35°C, and buffer containing 10 mM Mg^{2+} and 50 mM Na^{+}. (**E**) Ligation performed in the presence of target, with masking oligos, at 30°C, with buffer containing 10 mM Mg^{2+} and 50 mM Na^{+}. (**F**) Ligation performed in the presence of target, with masking oligos, at 30°C, with buffer containing 10 mM Mg^{2+} and 200 mM Na^{+}. All salts were chlorides. All solutions contained 20 mM Tris-HCI buffer (pH 7.8).

with the DNA target, and T4 DNA ligase was added to ligate probes that had hybridized to adjacent target sites. The ligated product was exponentially amplified by PCR using primers complementary to the constant regions of probes I and II (Fig. 1E, lane 2).

When we used the original experimental scheme shown in Figure 1A, we found that the hemirandom probes were ligated even in the absence of the DNA target (Fig. 1E, lane 1). This by-product was produced with much higher efficiency than anticipated. The target-independent ligation could arise from members of the library that serve as splints for hybridization and ligation of probes I and II, with random and constant regions being involved (Fig. 1B,C). To inhibit this process, we explored three different approaches. First, we decreased the size of the random regions of the probes to 7 nt and also used libraries of hemirandom probes having one fixed nucleotide position in the middle of random segments (Ho et al., 1996; Wrzesinski et al., 2000). However, we did not observe a significant reduction of the target-independent ligation with this approach. The second approach was to eliminate the self-ligating sequences by gel separation. In one version, nondenaturing gel electrophoresis was used to separate ligated hemirandom probes hybridized with the DNA target from nonhybridized probes before PCR amplification. In another version, negative selection was performed. T4 DNA ligase was added to the probes in the absence of target, and denaturing gel electrophoresis was used to separate probes that were not ligated from self-ligated probes. Surprisingly, we found that neither approach eliminated the target-independent ligation. Finally, we decided to use masking oligonucleotides (Liang et al., 2002; Paquin et al., 2000) complementary to the constant regions of the hemirandom probes, thus converting them from single-stranded to double-stranded from (Fig. 1D). The positive effect of the masking oligonucleotides on the reaction is seen in Figure 1E, lanes 3 and 4: target-independent ligation was strongly inhibited, and at the same time, target-dependent ligation was very efficient. With masking, we demonstrated target-dependent ligation over a range of sequences and lengths for the constant (20–26 nt) and random (12–7 nt) region of the probes and targets (from 20 nt to 7 kb) (data not shown). An additional benefit is that masking oligonucleotides prevent possible hybridization of the flanking constant regions of the probes with the target that could stabilize nonperfect complexes, thus allowing a number of mismatches. The efficiency of masking oligonucleotides in preventing target-independent ligation might be further improved by linking them to the constant regions of the hemirandom probes, forming terminal hairpin structures. To avoid problems in the PCR step due to potential blockage of primer binding by hairpin formation, additional restriction sites could be incorporated adjacent to the loops, so that after ligation, the loops can be removed by restriction enzymes, and the masking sequence can be melted off.

Length distribution of selected complementary sequences

Ligation of hemirandom probes was performed with heat-denatured cDNA of SFV and TNF-α DNA as targets. Ligation was performed at various temperatures (20°C–40°C) and salt concentrations (2–10 mM $MgCl_2$ and 50–200 mM NaCl), and the resulting products were cloned and sequenced to check for how well they matched the targets. Sequences of representative clones from ligation on the SFV target are shown in Figure 2. They are complementary to the sense and antisense strands since both target strands were present. The selected complementary regions typically contained sev-

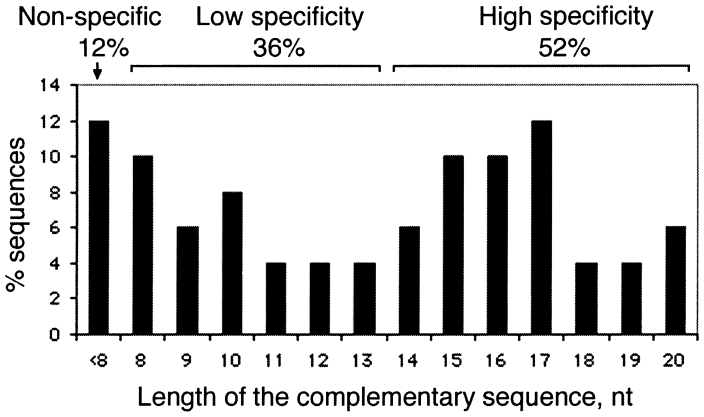

FIG. 3. Size distribution of target-matching SFV-specific library sequences. The diagram is based on the sequencing results of four independent ligation experiments performed at different temperature and salt concentrations, totaling 50 samples (see Fig. 2). For sequences longer than 10 nt, one mismatch was allowed.

eral mismatches, mostly distal to the ligation site but sometimes near it. Note that regions with target complementarity of ≤6 nt are always present in the samples (Fig. 2B) and are considered nonspecific and not statistically significant. The results are summarized in Figure 3. A relatively small number of the analyzed molecules have a near perfect 20-nt match with the target, and about 50% have 14–20 nt complementary to the target (allowing up to one mismatch). These lengths may provide enough specificity to recognize sequences that are unique in mammalian genomes (Saha et al., 2002).

The use of shorter randomized regions (e.g., 10 instead of 20 nt) in the hemirandom probes is beneficial because of the decreased stability of mismatched complexes. The known sensitivity of DNA ligase to mismatches adjacent to the ligation site additionally increases mismatch discrimination and reduces background false positives compared with approaches relying on hybridization alone (Gunderson et al., 1998; James et al., 1998; Landegren et al., 1988; Wu and Wallace, 1989). The fidelity of the directed library can be further improved by using more stringent hybridization conditions, including ligation at high salt concentration and elevated temperature (Barany, 1991). Indeed, we observed somewhat better results in reactions carried out at 30–35°C rather than 25°C. However, the temperature cannot be too high because the T_m for the 10-nt long oligonucleotides is in the range 20–40°C, depending on the sequence, and, thus, some AT-rich sequences can be lost or underrepresented in the library. Increase in Na^+ concentration (from 50 to 200 mM) and the addition of spermidine (5 mM) caused a modest increase in the length of the target-matching region in the library (Fig. 2 and Table 1) Variation in Mg^{2+} concentration was found to have little effect on the specificity of ligation.

In an attempt to increase the proportion of sequences matching the target, we studied the effect of adding short random oligoribonucleotides to the reaction. By providing competition with poorly matched hemirandom probes for target binding, we expected them to increase the average length of selected library sequences. Oligoribonucleotides 6 and 7 nt in length were found to have a negative effect (shorter complementary region), probably due to increasing the target-independent ligation of the probes by acting as ligation templates. Oligonucleotides 4 and 5 nt in length had no effect on the specificity of ligation (Table 1).

Yet another possibility for increasing the target specificity of the library is an affinity purification step that could also be included. For example, the DNA target can be chemically or enzymatically labeled with biotin. The target-dependent ligation products can then be separated from self-ligated probes by using avidin-containing beads and washing under stringent conditions.

Distribution of selected complementary sequences on DNA target

The distribution of the analyzed library sequences along both strands of SFV and SFV helper DNA is shown in Figure 4. Despite the limited number of sequences analyzed, this distribution is relatively even. The only exception is the region 7000–7378 near the end of the SFV genome, which is clearly overrepresented. However, when the TNF DNA target was provided in single-stranded form, the distribution was more even (not shown). In general, it is preferable to minimize the secondary structure of targets to increase the efficacy of their hybridization with the probes. However, it has been demonstrated in microarray experiments that neither ssDNA nor fragmented DNA has a clear advantage over dsDNA, as intramolecular pairing in ssDNA is not highly stable (Southern et al., 1999).

TABLE 1. SUMMARY OF SIZE DISTRIBUTION OF TARGET-MATCHING LIBRARY SEQUENCES, BASED ON RESULTS FROM LIGATION EXPERIMENTS USING HEAT-DENATURED SFV DNA AND SINGLE-STRANDED TNF DNA AS TEMPLATES

Target	*Probe random region, nt*	*Number of sequences*	*Nonspecific, % (< 8 nt)*	*Low specificity, % (8–13 nt)*	*High specificity, % (> 14 nt)*
SFV	10	50	12	36	52
TNF	10	38	16	42	42
TNF	12	21	24	38	38
TNF	7	14	28	29	43
TNF + spermidine	10	20	40	5	55
TNF + (rN) 4–5[a]	10	12	25	33	42
TNF + (rN) 6–7	10	12	67	33	0

[a]Short random oligonucleotides.

It should be noted that DNA ligases are also capable of ligating oligodeoxyribonucleotides on RNA templates, although with lower efficiency compared with DNA templates (Nilsson et al. 2001). Thus, a directed library can be produced by our method using RNA instead of DNA as target. However, because RNA is known to have stable intramolecular structures that can interfere with intermolecular hybridization, it should be fragmented, for example, by using alkaline hydrolysis (Southern et al., 1999). The use of RNA as a template also provides a method for directly mapping accessible sites on the RNA target. Hemirandom probes can be hybridized with a folded RNA target in vitro or in cell lysates and ligated, and the products can be amplified, cloned, and sequenced. As hybridization events occur only in the single-stranded and looped regions (and not in double-stranded stems), the method will provide information on accessible sites.

In conclusion, although for certain applications, conditions should be established that allow further reduction in the frequency of mismatches in the selected sequences, the ROLL method might find immediate biotechnologic application in the selection of target sites for hybridization (e.g., for antisense-based, siRNA-based, and ribozyme-based diagnostic probes, genomics tools, biosensors, imaging agents, and probes for blotting techniques, microarrays, and *in situ* hybridization) as well as first-pass selection of target sites for nucleic acid-based therapeutic agents. The method has a number of advantages over competing techniques: the scheme is very simple and fast, does not require complex manipulations and expensive enzymes, and allows variation in the length of the antisense sequence by varying the length of the random regions of the probes. Flexibility in use time- and cost-efficiency also make this method attractive for large-scale high-throughput applications.

ACKNOWLEDGMENTS

We gratefully acknowledge Prof. V. Vlassov (Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk 630090, Russia) and Dr. Anne Dallas (Somagenics, Inc.) for thoughtful discussions.

This research was supported by DARPA grant no. N65236-01-1-5437 and NIH grant no. 1R43GM 068225-01.

REFERENCES

ALLAWI, H.T., DONG, F., IP, H.S., NERI, B.P., and LYAMICHEV, V.I. (2001). Mapping of RNA accessible sites by extension of random oligonucleotide libraries with reverse transcriptase. RNA **7,** 314–327.

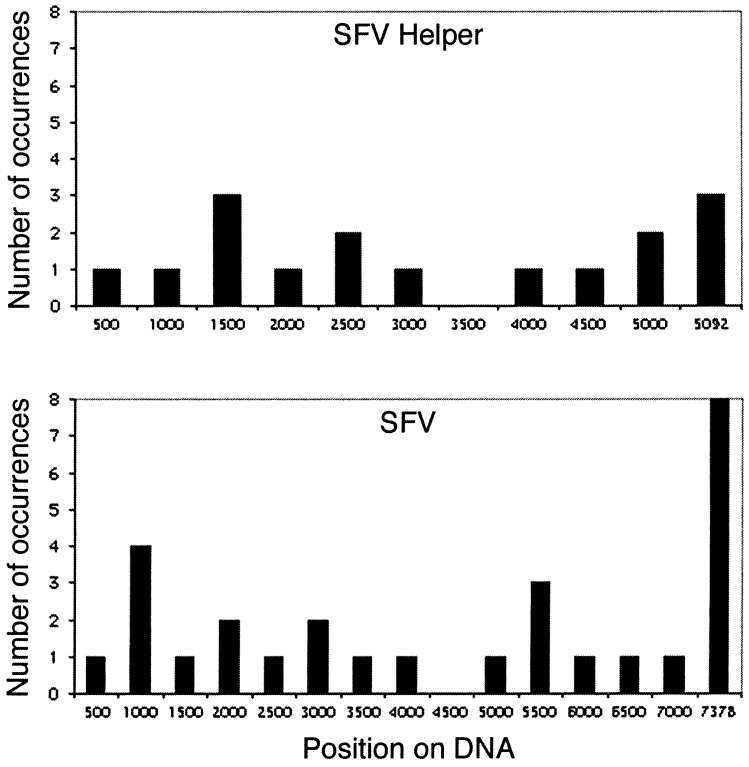

FIG. 4. Histogram showing the distribution of SFV-specific library sequences along the SFV DNA targets (replicon and helper fragments). Of 50 sequences determined, the 44 having more than 8 nt matching the target are shown.

AMARZGUIOUI, M., BREDE, G., BABAIE, E., GROTLI, M., SPROAT, B., and PRYDZ, H. (2000). Secondary structure prediction and *in vitro* accessibility of mRNA as tools in the selection of target sites for ribozymes. Nucleic Acids Res. **28,** 4113–4124.

BARANY, F. (1991). Genetic disease detection and DNA amplification using cloned thermostable ligase. Proc. Natl. Acad. Sci. USA **88,** 189–193.

BERTRAND, E.L., and ROSSI, J.J. (1994). Facilitation of hammerhead ribozyme catalysis by the nucleocapsid protein of HIV-1 and the heterogeneous nuclear ribonucleoprotein A1. EMBO J. **13,** 2904–2912.

BIRIKH, K.R., BERLIN, Y.A., SOREQ, H., and ECKSTEIN, F. (1997). Probing accessible sites for ribozymes on human acetylcholinesterase RNA. RNA **3,** 429–437.

BOIZIAU, C., DAUSSE, E., YURCHENKO, L., and TOULMÉ, J.J. (1999). DNA aptamers selected against the HIV-1 trans-activation-responsive RNA element from RNA-DNA kissing complexes. J. Biol. Chem. **274,** 12730–12737.

BRUICE, T.W., and LIMA, W.F. (1997). Control of complexity constraints on combinatorial screening for preferred oligonucleotide hybridization sites on structured RNA. Biochemistry **36,** 5004–5019.

BRUKNER, I., TREMBLAY, G.A., and PAQUIN, B. (2002). Generation of amplified genome-specific oligonucleotide probes and libraries. BioTechniques **33,** 874–881.

FAKLER, B., HERLITZE, S., AMTHOR, B., ZENNER, H.P., and RUPPERSBERG, J.P. (1994). Short antisense oligonucleotide-mediated inhibition is strongly dependent on oligo length and concentration but almost independent of location of the target sequence. J. Biol. Chem. **269,** 16187–16194.

FRANSEN, L., MULLER, R., MARMENOUT, A., TAVERNIER, J., VAN DER HEYDEN, J., KAWASHIMA, E., CHOLLET, A., TIZARD, R., VAN HEUVERSWYN, H., and VAN VLIET, A. (1985). Molecular cloning of mouse tumour necrosis factor cDNA and its eukaryotic expression. Nucleic Acids Res. **13,** 4417–4429.

GAROFF, H., FRISCHAUF, A.M., SIMONS, K., LEHRACH, H., and DELIUS, H. (1980). Nucleotide sequence of cDNA coding for Semliki Forest virus membrane glycoproteins. Nature **288,** 236–241.

GUNDERSON, K.L., HUANG, X.C., MORRIS, M.S., LIPSHUTZ, R.J., LOCKHART, D.J., and CHEE, M.S. (1998). Mutation detection by ligation to complete n-mer DNA arrays. Genome Res. **8,** 1142–1153.

HARADA, K., and ORGEL, L.E. (1993). *In vitro* selection of optimal DNA substrates for T4 RNA ligase. Proc. Natl. Acad. Sci. USA **90,** 1576–1579.

HO, S.P., BAO, Y., LESHER, T., MALHOTRA, R., MA, L.Y., FLUHARTY, S.J., and SAKAI, R.R. (1998). Mapping of RNA accessible sites for antisense experiments with oligonucleotide libraries. Nat. Biotechnol. **16,** 59–63.

HO, S.P., BRITTON, D.H., STONE, B.A., BEHRENS, D.L., LEFFET, L.M., HOBBS, F.W., MILLER, J.A., and TRAINOR, G.L. (1996). Potent antisense oligonucleotides to the human multidrug resistance-1 mRNA are rationally selected by mapping RNA-accessible sites with oligonucleotide libraries. Nucleic Acids Res. **24,** 1901–1907.

HOLEN, T., AMARZGUIOUI, M., WIIGER, M.T., BABAIE, E., and PRYDZ, H. (2002). Positional effects of short interfering RNAs targeting the human coagulation trigger tissue factor. Nucleic Acids Res. **30,** 1757–1766.

ISHIBASHI, O. (1997). A new method to synthesize competitor RNAs for accurate analyses by competitive RT-PCR. J. Biochem. Biophys. Methods **35,** 203–207.

JAMES, K., BOLES, A., HENCKEL, D., and ELLINGTON, A. (1998). The fidelity of template-directed oligonucleotide ligation and its relevance to DNA computation. Nucleic Acids Res. **26,** 5203–5211.

KAWASAKI, H., and TAIRA, K. (2002). A functional gene discovery in the Fas-mediated pathway to apoptosis by analysis of transiently expressed randomized hybrid-ribozyme libraries. Nucleic Acids Res. **30,** 3609–3614.

KRUGER, M., BEGER, C., LI, Q.X., WELCH, P.J., TRITZ, R., LEAVITT, M., BARBER, J.R., and WONG-STAAL, F. (2000). Identification of eIF2Bgamma and eIF2gamma as cofactors of hepatitis C virus internal ribosome entry site-mediated translation using a functional genomics approach. Proc. Natl. Acad. Sci. USA **97,** 8566–8571.

LANDEGREN, U., KAISER, R., SANDERS, J., and HOOD, L. (1988). A ligase-mediated gene detection technique. Science **241,** 1077–1080.

LANDEGREN, U., SAMIOTAKI, M., NILSSON, M., MALMGREN, H., and KWIATKOWSKI, M. (1996). Detecting genes with ligases. Methods **9,** 84–90.

LAPTEV, A.V., LU, Z., COLIGE, A., and PROCKOP, D.J. (1994). Specific inhibition of expression of a human collagen gene (COL1A1) with modified antisense oligonucleotides. The most effective target sites are clustered in double-stranded regions of the predicted secondary structure for the mRNA. Biochemistry **33,** 11033–11039.

LI, Q.X., ROBBINS, J.M., WELCH, P.J., WONG-STAAL, F., and BARBER, J.R. (2000). A novel functional genomics approach identifies mTERT as a suppressor of fibroblast transformation. Nucleic Acids Res. **28,** 2605–2612.

LIANG, Z., ZHANG, H.-Y., and WAHLESTEDT, C. (2002). Methods and means of RNA analysis. International Application (PCT) Patent No. WO 02/24950 A2.

LIEBER, A., and STRAUSS, M. (1995). Selection of efficient cleavage sites in target RNAs by using a ribozyme expression library. Mol. Cell. Biol. **15,** 540–551.

LIMA, W.F., BROWN-DRIVER, V., FOX, M., HANECAK, R., and BRUICE, T.W. (1997). Combinatorial screening and rational optimization for hybridization to folded hepatitis C virus RNA of oligonucleotides with biological antisense activity. J. Biol. Chem. **272,** 626–638.

LINDELL, M., ROMBY, P., and WAGNER, E.G. (2002). Lead(II) as a probe for investigating RNA structure *in vivo*. RNA **8,** 534–541.

LLOYD, B.H., GILES, R.V., SPILLER, D.G., GRZYBOWSKI, J., TIDD, D.M., and SIBSON, D.R. (2001). Determination of optimal sites of antisense oligonucleotide cleavage within TNFalpha mRNA. Nucleic Acids Res. **29,** 3664–3673.

MILNER, N., MIR, K.U., and SOUTHERN, E.M. (1997). Selecting effective antisense reagents on combinatorial oligonucleotide arrays. Nat. Biotechnol. **15,** 537–541.

NILSSON, M., ANTSON, D.O., BARBANY, G., and LANDEGREN, U. (2001). RNA-templated DNA ligation for transcript analysis. Nucleic Acids Res. **29,** 578–581.

PAQUIN, B., BRUKNER, I., and TREMBLAY, G. (2000).

Process for the generation of oligonucleotide libraries (OLs) representative of genomes or expressed mRNAs (cDNAs) and uses thereof. International Application (PCT) Patent No. WO 00/43538.

PATZEL, V., and SCZAKIEL, G. (2000). *In vitro* selection supports the view of a kinetic control of antisense RNA-mediated inhibition of gene expression in mammalian cells. Nucleic Acids Res. **28,** 2462–2466.

PENNICA, D., HAYFLICK, J.S., BRINGMAN, T.S., PALLADINO, M.A., and GOEDDEL, D.V. (1985). Cloning and expression in *Escherichia coli* of the cDNA for murine tumor necrosis factor. Proc. Natl. Acad. Sci. USA **82,** 6060–6064.

PIERCE, M.L., and RUFFNER, D.E. (1998). Construction of a directed hammerhead ribozyme library: Towards the identification of optimal target sites for antisense-mediated gene inhibition. Nucleic Acids Res. **26,** 5093–5101.

RUFFNER, D.E., PIERCE, M.L., and CHEN, Z. (1999). Directed antisense libraries. International Application (PCT) Patent No. WO99/50457.

SAHA, S., SPARKS, A.B., RAGO, C., AKMAEV, V., WANG, C.J., VOGELSTEIN, B., KINZLER, K.W., and VELCULESCU, V.E. (2002). Using the transcriptome to annotate the genome. Nat. Biotechnol. **20,** 508–512.

SCARABINO, D., and TOCCHINI-VALENTINI, G.P. (1996). Influence of substrate structure on cleavage by hammerhead ribozyme. FEBS Lett. **383,** 185–190.

SCHERR, M., LEBON, J., CASTANOTTO, D., CUNLIFFE, H.E., MELTZER, P.S., GANSER, A., RIGGS, A.D., and ROSSI, J.J. (2001). Detection of antisense and ribozyme accessible sites on native mRNAs: Application to NCOA3 mRNA. Mol. Ther. **4,** 454–460.

SCZAKIEL, G., and FAR, R.K. (2002). The role of target accessibility for antisense inhibition. Curr. Opin. Mol. Ther. **4,** 149–153.

SEN, G., WEHRMAN, T.S., MYERS, J.W., and BLAU, H.M. (2004). Restriction enzyme-generated siRNA (REGS) vectors and libraries. Nat. Genet. **36,** 183–189.

SHIRANE, D., SUGAO, K., NAMIKI, S., TANABE, M., IINO, M., and HIROSE, K. (2004). Enzymatic production of RNAi libraries from cDNAs. Nat. Genet. **36,** 190–196.

SOHAIL, M., and SOUTHERN, E.M. (2000). Selecting optimal antisense reagents. Adv. Drug Deliv. Rev. **44,** 23–34.

SOUTHERN, E., MIR, K., and SHCHEPINOV, M. (1999). Molecular interactions on microarrays. Nat. Genet. **21,** 5–9.

TAKKINEN, K. (1986). Complete nucleotide sequence of the nonstructural protein genes of Semliki Forest virus. Nucleic Acids Res. **14,** 5667–5682.

WRZESINSKI, J., LEGIEWICZ, M., and CIESIOLKA, J. (2000). Mapping of accessible sites for oligonucleotide hybridization on hepatitis delta virus ribozymes. Nucleic Acids Res. **28,** 1785–1793.

WU, D.Y., and WALLACE, R.B. (1989). The ligation amplification reaction (LAR)—Amplification of specific DNA sequences using sequential rounds of template-dependent ligation. Genomics **4,** 560–569.

YU, Q., PECCHIA, D.B., KINGSLEY, S.L., HECKMAN, J.E., and BURKE, J.M. (1998). Cleavage of highly structured viral RNA molecules by combinatorial libraries of hairpin ribozymes. The most effective ribozymes are not predicted by substrate selection rules. J. Biol. Chem. **273,** 23524–23533.

Address reprint requests to:
Dr. Alexander V. Vlassov
SomaGenics, Inc.
2161 Delaware Avenue
Santa Cruz, CA 95060

E-mail: avlassov@somagenics.com

or

Dr. Sergei A. Kazakov
SomaGenics, Inc.
2161 Delaware Avenue
Santa Cruz, CA 95060

E-mail: skazakov@somagenics.com

Received May 24, 2004; accepted in revised form June 23, 2004.