mRNA accessible site tagging (MAST): a novel high throughput method for selecting effective antisense oligonucleotides

Hong-Yan Zhang1,2, Jianping Mao1, Daixing Zhou1, Yunhe Xu1, Håkan Thonberg1, Zicai Liang1,* and Claes Wahlestedt1

1Center for Genomics and Bioinformatics, Karolinska Institutet, 17177, Stockholm, Sweden and 2United Biotechnologies AB, Sevensvägen 4, 17446, Sundbyberg, Sweden

Received March 5, 2003; Revised and Accepted May 8, 2003

ABSTRACT

A solution-based method, mRNA accessible site tagging (MAST), has been developed to map the accessible sites of any given mRNA in high throughput fashion. mRNA molecules were immobilized and hybridized to randomized oligonucleotide libraries. Oligonucleotides specifically hybridized to the mRNA were sequenced and found to be able to precisely define the accessible sites of the mRNA. A number of ways were used to validate the accessible sites defined by the MAST process. Mapping of rabbit β-globin mRNA demonstrates the efficacy and advantage of MAST over other technologies in identifying accessible sites. Antisense oligonucleotides designed according to the accessible site map of human RhoA and Renilla luciferase mRNA result in knockdown effects that are in good correlation with the degrees of accessibility. The MAST methodology can be applied to mRNA of any length using a universal protocol.

INTRODUCTION

Antisense technology is a powerful tool for functional genomics and drug discovery. By pairing with their cognate mRNA, antisense oligonucleotides can knockdown gene expression specifically through mechanisms such as RNase H-mediated mRNA cleavage, destabilization of the target mRNA or aberration of translation or splicing (1–3). The antisense oligonucleotide itself can also be an effective drug. Because of the economical value and short experimental cycle, antisense technology has been widely accepted as the tool to study functions of a gene and to validate drug targets (4). Although the small interfering RNAs (siRNAs) as a new gene knockdown technology have shown significant advantages in some applications (5), antisense reagents still remain as a competitive alternative for functional gene validation and drug development (6).

Selection of antisense sequences represents an important step in the application of antisense technology. Empirically, only a small proportion of the tested antisense oligonucleotides (normally 2–10%) exhibit effective sequence-specific antisense activities. It is well recognized that the effectiveness of an antisense oligonucleotide depends on the accessibility of the mRNA site to which it is intended to bind. More effort is normally needed to experimentally characterize the efficacy, specificity and other properties of antisense oligonucleotides after the oligonucleotides have been chosen, but accurate identification of accessible sites can dramatically reduce the downstream workload. Methods developed for identification of accessible sites include RNase H mapping (7,8), gel shift (9), oligonucleotide array (10) and random RT priming (11), as well as computational predictions (12–14). Among them, RNase H mapping and random RT priming use short oligonucleotide libraries to determine the accessibility of a mRNA and have shown promising results. However, both of them have to use high resolution gels, which have very limited throughput, to define the cleavage or priming sites. Microarrays containing all the oligonucleotides that cover a mRNA sequence have been used to identify the accessible sites of the mRNA by hybridization with good efficacy (10). Such microarrays have to be individually designed and manufactured for each different mRNA analyzed. This imposes a very high cost and serious restrictions on the throughput, which in turn have prevented this great method from being widely used. Due to the practical complexities or poor availability of the existing experimental methods for identifying accessible sites, computational approaches have also been attempted for making in silico predictions (12–14). Applicability of these prediction models beyond the training set of genes is, however, still questionable.

The number of genes in the human genome is predicted to be in the range 30 000–50 000. The majority of them are genes with unknown functions. It is economically plausible to use technologies such as antisense to functionally characterize these genes. Throughput-wise, however, none of the existing experimental approaches can offer the mapping capacity to resolve the mRNA accessibility of a fraction of these genes. In
this report, a simple bench-top method, called MAST, has been designed to provide high throughput mapping of mRNA accessibility at high precision using standard molecular biology procedures.

MATERIALS AND METHODS

Reagents

Oligonucleotides were purchased from Interactiva and MWG, Germany. The DYEnamic terminator sequencing kit was from Amersham Biosciences, Sweden. pGEM-T vector and competent Escherichia coli were from Promega, USA. Streptavidin-coated paramagnetic beads (Dynabeads) were purchased from Dynal, Norway. DEPC was from Sigma, USA. The PCR purification kit was from Qiagen, USA.

Rabbit β-globin cDNA was RT–PCR amplified directly from rabbit globin mRNA purchased from Life Technologies AB, Sweden.

The combinatorial libraries

In each library, three oligonucleotides were synthesized separately. The library contains 18 totally randomized nucleotides nested in two stretches of clamping sequences (on 5' end, TGC AGA CAG TAA AGG GAT CT; on 3' end, CGG CCG) (Fig. 1). The clamping sequences were made double-stranded by annealing to equimolar amounts (100 µM each) of blocking oligonucleotides (AGA TCC TTA TAC CTG CA and CGC CCG) in 2× SSC (300 mM NaCl, 50 mM sodium citrate, pH 7) using a temperature touchdown program (94°C for 3 min and then 92°C for 20 s, 90°C for 20 s, 88°C for 20 s and so on to 30°C for 20 s). The library was then stored at 4°C. This modification left only the randomized region available for subsequent hybridization steps under non-denaturing conditions at 37–40°C. The 3' clamping sequence was made small so that it would afford minimal steric hindrance and give more flexibility to the randomized region of the library during hybridization to the mRNA. A short 3' clamping sequence in the standard libraries has enabled the easy amplification of selected MAST tags (see below), but on a few occasions it was noticed that this supposedly double-stranded region might also have taken part in hybridization to the mRNA target.

Two strategies have been designed for high throughput analysis of the tag sequences. Firstly, concatemerization of the tags was performed using the following procedure. Singly amplified tags were cleaved with EcoRI and BamHI. The tag fragments nested between the two sites were resolved on 20% polyacrylamide gels (acrylamide:bis-acrylamide 30:1). The DNA was cut out from the gel and eluted into 300 mM sodium acetate buffer (pH 5.2) overnight and precipitated together with 1 µl Glycolblue (15 mg/ml). The fragments were then resuspended in 20 µl H2O and ligated with T4 ligase for 2 h at room temperature. The ligation products were size selected on a 1% agarose gel to recover fragments that were in the range 250–600 bp (data not shown).

In light of the significant manual handling in the concatemerization process, a second strategy based on the pyrosequencing method has been extensively tested and amended for a high degree of automation. PCR-amplified tags were directly subcloned into pGEM-T vectors. After overnight incubation, single bacterial colonies were picked with a BioPick robot (Tican, Germany) and the tag from each colony was PCR amplified, immobilized onto magnetic beads and sequenced on a Pyrosequencer (Pyrosequencing AB, Sweden) according to the manufacturer’s instructions.

Preparation of biotin-labeled mRNA

cDNA fragments were tagged with T7, T3 or Sp6 promoters during PCR amplification and were used to produce the corresponding mRNA (cRNA) by in vitro transcription reactions driven by T7, T3 or Sp6 RNA polymerase according to the procedures of the manufacturers, except that all transcription reactions were supplemented with 0.1 mM biotin-UTP (Amersham Pharmacia Biotech, Sweden) in addition to 1 mM each ATP, UTP, CTP and GTP. The products were normally analyzed in a 1% agarose gel to control for the quality of mRNA.

RNA immobilization and oligonucleotide selection

Typically, 100 µl of suspended Dynabeads was washed with 200 µl of DEPC-treated 2× SSC 10 times. Then the beads were resuspended in 50 µl of 5× SSC containing 5 µg biotin-labeled mRNA and the binding reaction was allowed to proceed for 30 min at room temperature with constant shaking. Afterwards beads loaded with mRNA were washed five times in 2× SSC and five times with 2× SST (SSC solution supplemented with 0.1% Tween-20). The beads were then resuspended in 100 µl of 2× SST containing 1 µl of the combinatorial library and hybridization of the immobilized mRNA and the oligonucleotides from the library was allowed to proceed for 1 h at 40°C with constant shaking. The beads were sequentially washed with 2× SST 10 times at 40°C and 1× SST five times at room temperature and rinsed five times with 1× SSC. Then the beads were resuspended in 50 µl of H2O and boiled for 2 min and the bound oligonucleotides were recovered in the aqueous phase. The recovered oligonucleotides will be referred to as accessible site tags or tags in this paper.

Amplification and sequencing of accessible site tags

One microliter of recovered oligonucleotides was PCR amplified using a forward primer (GAT GCA GAC AGT AAA GGG AT) that overlaps with the 5' clamping sequence and a reverse primer (CAA CTA AGG GAA TCC GCC CG) that partially complements the 3' clamping sequence. The PCR was carried out firstly for one cycle of 94°C for 2 min, 37°C for 1 min, 40°C for 1 min, 45°C for 1 min, 72°C for 1 min to allow the generation of full priming sites, and then 30 cycles of 94°C for 30 s, 50°C for 30 s and 72°C for 30 s. Taq polymerase and buffer were purchased from Promega, USA. PCR products (amplified tags) or concatemerized tags were purified and subcloned into pGEM-T vectors. Sequencing was done either by using the DYEnamic terminator sequencing kit according to the manufacturer’s instructions or by pyrosequencing according to the instructions from Pyrosequencing AB, Sweden. Alignment of the oligonucleotide sequences to the target gene was accomplished by DNAStrider™ software or RiboAlign™, software designed by Neuromics Inc., USA.
Knockdown of the endogenous RhoA gene was achieved by transfecting phosphothioate antisense oligonucleotides into human embryonic kidney cells (HEK 293 cells). Cells were grown at 37°C, 5% CO₂ in Dulbecco’s modified Eagle’s medium (Gibco), seeded in 24-well plates at 60% confluence and transfected on the second day. Antisense oligonucleotides (10 μg/well, equivalent to 330 nM final concentration) (Annovis Inc., Aston, PA) were used to treat the cells in the presence of Lipofectamine 2000 from Invitrogen Inc. (Carlsbad, CA) according to the manufacturer’s protocol. Twelve hours after the treatment, total RNA was purified from the cells using a RNAeasy kit from Qiagen (Valencia, CA) and the levels of RhoA and GAPDH RNA were analyzed by real time PCR using an ABI7700 Sequence Detector from Applied Biosystems Group (Foster City, CA). Oligonucleotides targeting three accessible sites and 10 other sites (Table 1) with various levels of accessibility were tested for their accessibility. Randomized oligonucleotides

**Figure 1.** Schematic diagram of the MAST method. (A) An oligonucleotide library with an 18 nt stretch of fully randomized sequence nested between two stretches of clamping sequences (on 5’ end, TGC AGA CAG TAA AGG GAT CT; on 3’ end, CGG GCG). The clamping sequences were blocked by complementary sequences in the library. They were designed to facilitate subsequent PCR amplification of the library while not interfering with the hybridization of the single-stranded region. (B) Biotin-labeled mRNA was synthesized by *in vitro* transcription and bound to streptavidin-coated paramagnetic beads. The oligonucleotide library was then allowed to hybridize with the immobilized mRNA under controlled temperature and salt concentration. After unbound and non-specifically bound oligonucleotides were removed by washing at proper stringency, oligonucleotides specifically bound to the mRNA were eluted by boiling in water. Then the oligonucleotides (ASTs) were PCR amplified using GAT GCA GAC AGT AAA GGG AT and CAA CTA AGG GAA TTC GCC CG as primers. The PCR products were cloned into vectors for normal or high throughput sequencing. The sequences of those oligonucleotides were then used *in silico* to define the accessible sites.
and Lipofectamine 2000 were used as the control. The relative RNA level of RhoA was normalized against that of GAPDH. Experiments were repeated three times with identical results.

**Functional analysis of the Renilla luciferase antisense oligonucleotides**

*Renilla* luciferase (accession no. AF025846) mRNA was transcribed in vitro from a plasmid (Promega) and mapped using the procedure described above. Cells were maintained as described for the RhoA experiments. The HEK 293 cells were co-transfected with plasmids for both *Renilla* and firefly luciferases and different concentrations of antisense DNA in the presence of Lipofectamine 2000 (Invitrogen, USA) at a final concentration of 0.17% in the total transfection volume. For each well of HEK 293 cells, 0.17 g of pRL-TK were used. The transfection medium was changed to culture medium (1 ml) after 4 h of co-transfection. The cells were collected 22 h post-transfection, prepared using the procedure described above. Cells were maintained as in vitro transcribed luciferase (accession no. AF025846) mRNA was eluted in H2O at 100 °C for 2 min, PCR amplified and sequenced. After alignment to β-globin mRNA, 13 of 19 ASTs sequenced pinpointed two regions of the 122 nt β-globin mRNA fragment: nucleotides 40–62 (region A) and 67–85 (region B) (Fig. 2). The identification of region B as an accessible site is in very good agreement with results from the scan array data (10). The identification of region B also nicely matched with the array based results qualitatively, but differences emerged when the yields of AST (or hybridization yield in the array method), presumably the degrees of accessibility, were compared. A similar number of ASTs was recovered for region B as compared with region A, whereas in the array method region B gave several fold lower hybridization yield. From the counts of the ASTs, region B appears to be almost as accessible as region A, whereas region B was defined as a marginally accessible site by the scan array method. Secondary structure modeling suggested that region B is predominantly a single-stranded region, with no significant intramolecular double helix formation. This seems to be in better agreement with the predictions of the MAST method than of the array method.

Interestingly, it is observed that opposite mapping results have been generated by the RT priming and oligonucleotide array methods (Fig. 2B). Using the RT priming method, the strongest accessible site was determined to be region B, whereas region A only appeared to be marginally accessible (11). It is not known how this discrepancy could have arisen. It was speculated that the two methods rely on different mechanisms of accessible site detection (11). Under the experimental conditions used in their respective assays, the array-based method is more likely to detect the formation of very stable DNA–RNA heteroduplex, whereas the RT priming method is more likely to identify mRNA sites that can accommodate the quickly hybridizing oligonucleotide species. The unambiguous identification of both region A and region B of β-globin by MAST under uniform experimental conditions demonstrated that the MAST process might possess the combined advantages of the RT priming and array methods in terms of simultaneous detection of both fast hybridizing and stably hybridizing accessible sites.

**MASH mapping of longer mRNA targets**

The average size of a mRNA molecule is ~1.5 kb. A scan array for mapping such a mRNA over its entire length needs about 1500 cycles (deprotection and coupling) of chemical synthesis, which means enormous consumption of time and reagents. Similarly, all gel-based methods of accessibility mapping are limited by the resolving power of the acrylamide gels.

---

**Table 1. Sequences of antisense oligonucleotides tested for inhibition of endogenous RhoA mRNA level in HEK 293 cells**

<table>
<thead>
<tr>
<th>Name</th>
<th>Site</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site A</td>
<td>17–34</td>
<td>AGCTGAGGACCGAGCCGT</td>
</tr>
<tr>
<td>Site B</td>
<td>256–273</td>
<td>CTGCTTTCCTACACACTTC</td>
</tr>
<tr>
<td>Site C</td>
<td>536–553</td>
<td>CTCGCCTTCTTCTAGGT</td>
</tr>
<tr>
<td>CTL 1</td>
<td>1–18</td>
<td>GTGACGTACAGAGACAGAC</td>
</tr>
<tr>
<td>CTL 2</td>
<td>5–22</td>
<td>GACCDTGACCTAAGGAGGA</td>
</tr>
<tr>
<td>CTL 3</td>
<td>423–440</td>
<td>GGCAAGAAATCTTCGCT</td>
</tr>
<tr>
<td>CTL 4</td>
<td>433–451</td>
<td>TGGGACCTTGAGCAGAAG</td>
</tr>
<tr>
<td>CTL 5</td>
<td>465–482</td>
<td>CGAAGATCCTTCTTATTC</td>
</tr>
<tr>
<td>CTL 6</td>
<td>542–563</td>
<td>TTCCCATATCTCTGCTCTTT</td>
</tr>
<tr>
<td>CTL 7</td>
<td>599–616</td>
<td>CCTCATCTTGCTCTTGG</td>
</tr>
<tr>
<td>CTL 8</td>
<td>963–982</td>
<td>AAAGAACGAAGAAGTTAAGA</td>
</tr>
<tr>
<td>CTL 9</td>
<td>1080–1098</td>
<td>CTGAAATGGTGAATCC</td>
</tr>
<tr>
<td>CTL 10</td>
<td>1281–1299</td>
<td>CTAGCTGGGGTAGATATAT</td>
</tr>
</tbody>
</table>

Sites A, B and C were the oligonucleotides targeting the three accessible sites and CTL 1–CTL 10 were control oligonucleotides against RhoA mRNA sites with low accessibility. *Sites were numbered according to mRNA sequence presented in GenBank entry BC005976. All antisense oligonucleotides were synthesized as phosphorothioates.*
Therefore, a long mRNA will have to be mapped in fragments of ~0.5 kb length. The complexity of the procedures comprises the major obstacle that prevents possible transformation of these methods into a high throughput mapping scheme. Using MAST we were able to map mRNA of any length with a single robust protocol and to generate accessibility maps for the mRNA (Fig. 3A). We have successfully mapped mRNA with lengths from 0.1 up to 6 kb.

Activity assay of antisense sequences defined by the MAST method

It has been proposed that the efficacy of an antisense oligonucleotide is dependent on the accessibility of the sequence (8,11). To test the hypothesis that an accessible site defined by the MAST process actually presents a potential effective antisense sequence, we tested antisense activity on the human RhoA gene, an endogenously expressed gene in HEK 293 cells.

The graphic form of the accessible map is shown in Figure 3A, where the frequency of the tag alignment to each location of the mRNA is demonstrated in color code. When ranking the accessibility of different sites with the aligned tags, the quality of tags (in terms of length of alignment, gaps and T-G matches) was also taken into consideration. With both tag frequency and quality considered, MAST mapping of human RhoA mRNA identified three predominant accessible sites. We then designed three antisense oligonucleotides against the accessible sites based on the map and analyzed their antisense activities in HEK 293 cells by measuring RhoA mRNA levels using the Taqman assay. As a comparison, 10 additional oligonucleotides targeting RhoA mRNA sites with medium, low and no accessibility were included in the test. All three antisense oligonucleotides targeting accessible sites were found to knockdown the expression of the mRNA by >80% at a concentration of 330 nM, whereas sites with medium and low accessibility showed much lower levels of
inhibition (Fig. 3B). This was in contrast to previous results where only 1 out of 38 empirically selected antisense oligonucleotides reached this level of inhibition (15). In this sense, we observed excellent correlations between the efficacy of an antisense oligonucleotide and the accessibility of the sequence. CTL 10 was not identified as an accessible site but it did significantly reduce the RhoA mRNA level. We have two possible explanations for this exception. Firstly, it might have been caused by minor structural changes induced by other factors in the cell since this site is very close to the poly(A) tail. Secondly, we might simply have missed this site because we did not select enough tags to cover the mRNA. Since CTL 10 was not identified as an accessible site in repeated mapping, it is more likely that the activity of CTL 10 resulted from interactions between the mRNA and some cellular components.

CTL 6 and Site C share most of their sequences except for 6 nt, but the activity of CTL 6 was nearly half of that observed for the Site C antisense oligonucleotide. We attributed this sharp difference in activity to the fact that these 6 nt are an essential part of this strong accessible site. Similar observations were made for CTL 1, CTL 2 and Site A. Among these three partially overlapping oligonucleotides, the one designed against the accessible site (Site A) gave the strongest.

A.
knockdown. This attested to the belief that a precise definition of the accessible site is very important in achieving optimal antisense inhibition on mRNA targets. The results from the mapping of about 961 genes suggested that there are about one to three highly accessible sites per kilobase of mRNA.

The luciferase assay has been widely used for the demonstration of siRNA-based gene inhibition, but there have been no data available for antisense accessibility in either Renilla luciferase mRNA or firefly luciferase mRNA apart from a recent empirical antisense study of firefly luciferase using locked nucleic acids (16). We decided to map the Renilla luciferase mRNA for its antisense accessibility and compared the dose–response curves of accessible and non-accessible sites. The 935 nt mRNA was mapped using the standard protocol as described in Materials and Methods and one predominant accessible site was identified between 20 and 40 nt (Fig. 4A). An antisense oligonucleotide (RLas21–38, CGT TTC CTT TGT TCT GGA) targeting this accessible site and three control antisense oligonucleotides (RLas375–392, TAG CTA TAA TGA AAT GCC; RLas502, TTT TTC TCC TTC TTC AGA; RLas614, CAC CTT TCT CTT TGA ATG) targeting non-accessible sites on the mRNA were synthesized with phosphothioate modifications and tested for antisense activity. The antisense oligonucleotide (RLas21–38) targeting this accessible site showed high inhibitory activity and knocked down the expression of Renilla luciferase by 86% at 26 nM, whereas the other oligonucleotides targeting non-accessible sites did not show significant activity (36, 26 and 41% inhibition, respectively). The inhibition induced by RLas21–38 was nicely correlated with the dose of the oligonucleotide (Fig. 4B). While antisense oligonucleotides binding to other regions of the mRNA are believed to rely mostly on an RNase H-based mechanism to achieve inhibitory effects, antisense oligonucleotides binding to the vicinity of the translation start codon can also inhibit expression by physical blocking (steric hindrance) of the translation apparatus. In eukaryotic cells such blockade can only be broadly achieved by high affinity DNA analogs such as peptide nucleic acids, locked nucleic acids and morpholino nucleic acids. It was demonstrated that such molecules do not work efficiently if they are used to target sites 20 nt or more downstream of the AUG. Given that we are using a normal phosphorothioate oligonucleotide 21 nt downstream of the AUG, the observed antisense activity of AS21–38 should have been achieved through an RNase H mechanism, thus being accessibility dependent.

In summary, empirical testing and experimental assays are widely used to predict effective antisense sequence and all existing experimental methods suffer from cumbersome procedures and low throughput. The MAST method described here is simple and easy to perform in any biology laboratory. Thorough interrogation of a small number of mRNAs (10 mRNAs, for example) can be accomplished within 3 days using this method. This throughput level should meet...
most laboratory needs in terms of antisense sequence selection. Unlike other methods, the MAST procedure is designed to be scaled up. No adjustment is needed in the MAST procedure when performing analysis on multiple mRNAs or mRNAs of different lengths. It is theoretically possible to use this method to investigate tens to hundreds of mRNAs in the same reaction tube, thus affording unrivalled parallel processing capacity in mRNA accessibility analysis. We have successfully mapped accessible sites for more than 1000 genes within 3 months. Apart from using this method to select antisense oligonucleotides for routine work, one potential use of this technology is to build an mRNA accessibility database by MAST mapping of all known human (or any other species) mRNAs.

ACKNOWLEDGEMENTS
This work was supported by grants from VINNOVA (the Swedish Agency for Innovation Systems) to Z.L. and from Pharmacia Corp. to Z.L. and C.W. H.-Y.Z. was partially supported by a repatriation grant from Naturvetenskapliga Forskningsrådet. United Biotechnologies AB is acknowledged for contributions to past and continued development of the MAST technology.
REFERENCES