

Direct labeling of RNA with multiple biotins allows sensitive expression profiling of acute leukemia class predictor genes

Kyle Cole*, Vivi Truong, Dale Barone and Glenn McGall

Affymetrix Inc., 3380 Central Expressway, Santa Clara, CA 95051, USA

Received November 26, 2003; Revised January 20, 2004; Accepted May 26, 2004

ABSTRACT

Direct labeling of RNA is an expedient method for labeling large quantities (e.g. micrograms) of target RNA for microarray analysis. We have developed an efficient labeling system that uses T4 RNA ligase to attach a 3'-biotinylated donor molecule to target RNA. Microarray analyses indicate that directly labeled RNA is uniformly labeled, has higher signal intensity than comparable labeling methods and achieves high transcript detection sensitivity. The labeled donor molecule we have developed allows the attachment of multiple biotins, which increases target signal intensity up to 30%. We have used this direct-labeling method to detect previously discovered class predictor genes for two types of cancer: acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). In order to test the sensitivity of direct RNA labeling, we analyzed the AML and ALL expression profiles for predictor genes that were previously found to show elevated expression in the disease state. Direct labeling of AML poly(A) RNA detects 90% of the class predictor genes that are detected by the IVT-based target amplification method used to discover the genes. These results indicate that the detection sensitivity, simplicity (single tube reaction) and speed (2 h) of this direct labeling protocol may be ideal for diagnostic applications that do not require target amplification.

INTRODUCTION

DNA microarrays have revolutionized gene expression profiling by allowing highly parallel and quantitative monitoring of specific transcripts. Despite this extensive profiling capability, the use of microarrays for clinical diagnostics is not yet prevalent. One factor limiting microarray use is the cost and complexity of target preparation. Many current methods of sample preparation rely on several enzymatic steps to copy, amplify and label nucleic acid target (e.g. reverse transcription, *in vitro* transcription or PCR). Although many of these methods effectively detect low-abundance mRNAs, representation of the initial transcript population may be skewed by enzymatic amplification (1,2). In contrast, direct labeling of

RNA does not involve amplification and requires fewer enzymatic manipulations, thus accurately preserving relative transcript abundance and simplifying target preparation. Direct labeling of mRNA is an expedient alternative for microarray applications that do not require extremely high detection sensitivity. In this paper, we report a direct-labeling method that uses T4 RNA ligase to attach a biotinylated nucleotide to the 3' end of RNA targets. This method labels RNA fragments uniformly, thus avoiding sequence bias associated with the incorporation of labeled nucleotides during synthesis (e.g. biotin-dCTP). In contrast to other direct-labeling methods that label the nucleic acid strand internally, such as biotin-ULS (3,4), or chemically modify the nucleobase (5), end-labeled RNA is predicted to have higher target-probe affinity because hybridization is unimpaired by label moieties (6). Because of the 3'→5' orientation of probes on the microarray surface, the 3' position of the label should expose the biotin for efficient binding by the streptavidin-fluorophore.

T4 RNA ligase catalyzes the 3'→5' phosphodiester bond formation of RNA molecules utilizing the hydrolysis of ATP to AMP and PPi (7). The direct-labeling system we have developed uses T4 RNA ligase to attach a 3'-biotinylated nucleotide donor to the 3'-hydroxyl of an RNA acceptor. Although T4 RNA ligase has previously been used to label RNA with various moieties [e.g. radioactive nucleotides, fluorophores and amino acids (8–10)], this is the first report that we are aware of that uses T4 RNA ligase to label a heterogeneous RNA sample for microarray analysis. One challenge in adapting this labeling system for use with microarrays is that RNA ligation efficiency can vary significantly depending on the type of donor molecule used and the acceptor size and sequence (11–14). Alternative enzymatic methods for end-labeling RNA also suffer from similar inefficiencies (15,16). In this paper, we optimize the conditions for RNA ligation and describe an enzymatic fragmentation method that generates RNA fragments that are optimal for end-labeling and the correct size for hybridization to DNA microarrays.

The nucleotide donors used in this study contain biotin moieties tethered to the 3'-hydroxyl, rather than attached to the nucleobase. Attachment at this position has the advantage of allowing multiple biotin molecules to be affixed sequentially to the donor nucleotide without reducing ligation efficiency. End-labeling RNA target with multiple biotins has the potential to significantly enhance overall signal intensity (6) and improve the detection of low-abundance transcripts.

*To whom correspondence should be addressed. Tel: +1 408 731 5776; Fax: +1 408 481 0422; Email: kyle_cole@affymetrix.com

The simplicity and cost-effectiveness of direct labeling may be valuable for many clinical diagnostic applications that require robust and inexpensive assay protocols, such as cancer classification. In order to demonstrate the feasibility of direct labeling for leukemia classification we determined the gene expression profiles of two leukemia types, acute myeloid leukemia (AML, KG-1 cell line) and acute lymphoblastic leukemia (ALL, MOLT-4 cell line). Expression profiles for AML total RNA and poly(A) RNA and for ALL poly(A) RNA were analyzed for the detection of subclass predictor genes identified in previous microarray studies (17,18). Our results indicate that many class predictor genes could be detected from both total RNA and poly(A) RNA for both leukemia types, suggesting that direct labeling may be a viable method for certain diagnostic applications.

MATERIALS AND METHODS

Synthesis of donor molecules

The biotinylated donor molecules used in this study (see Table 1) were synthesized on an ABI 394 DNA synthesizer using standard synthesis protocols and phosphoramidite reagents and supports from Glen Research (Figure 1). For example, starting with a 3'-biotin-TEG CPG support, 0–4 sequential additions of hexaethyleneglycol (HEG)

Table 1. Donor molecules for RNA labeling

Nomenclature	Compound	Relative ligation efficiency
pApB	5'-pAp-TEG-biotin-3'	0.9 ^a
pA ₅ pB	5'-pA ₅ p-TEG-biotin-3'	0.5 ^a
pCpB	5'-pCp-TEG-biotin-3'	1 ^{ab}
AppCpB	A(5')pp(5')Cp-TEG-biotin-3'	1 ^a
pCpB ₂	5'-pCp-(HEGTEG-biotin) ₂ -3'	n.d.
pCpB ₃	5'-pCp-(HEGTEG-biotin) ₃ -3'	n.d.
pCpB ₅	5'-pCp-(HEGTEG-biotin) ₅ -3'	0.8 ^b

n.d., not determined.

^aBased on array intensities of cRNA target.

^bBased on gel shift analysis or RNA 20mer acceptor.

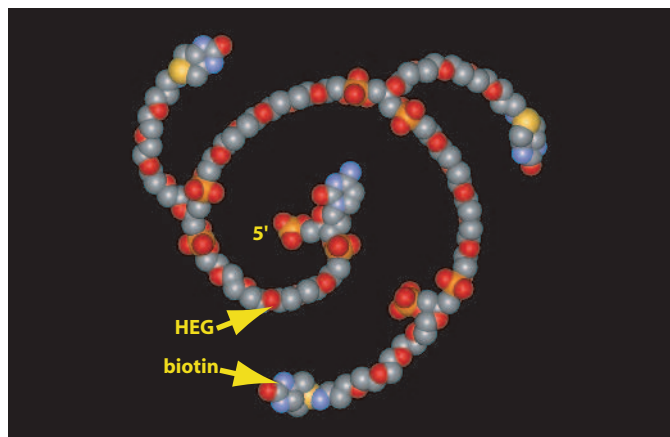


Figure 1. Structure of nucleotide donor molecule pCpB₃. Multiple biotin labels attached to the donor 3'-hydroxyl are separated by HEG and TEG spacers. Concatenation of at least five biotins to the donor molecule can be accomplished without significantly inhibiting ligation efficiency.

phosphoramidite followed by biotin-TEG phosphoramidite were performed. This was followed by the addition of C phosphoramidite, and finally 5'-phosphate-ON. After removing the protecting groups using recommended procedures, the products were purified by reverse-phase high-performance liquid chromatography (HPLC) to >90% purity and characterized by either ESI or MALDI-TOF mass spectroscopy.

The pre-adenylated pyrophosphate donor, A(5')pp(5')Cp (biotin-TEG)-3', was prepared by solution-phase condensation (Figure 2) of p(5')Cp(biotin-TEG)-3' and adenosine-5'-monophosphoromorpholidate (Sigma), according to literature procedures (7,16–18). The product was purified by reverse-phase followed by ion-exchange HPLC, and then characterized by MALDI-TOF MS.

Target RNA preparation

Unlabeled and internally labeled target cRNA was generated following the standard Affymetrix protocol for eukaryotic gene expression analysis, except as noted. Briefly, 10 µg of total human heart RNA (Ambion) or HeLa total RNA were reverse transcribed using a T7-dT₂₄ primer and converted into double-stranded cDNA containing a T7 promoter (SuperScript II RT kit, Invitrogen). Internally labeled cRNA was transcribed from this cDNA template using either a T7 Megascript kit (Ambion) with the addition of biotin-CTP and biotin-UTP (NEN) or a BioArray High Yield RNA Transcript labeling kit (Enzo). cRNA was then purified using RNeasy columns (Qiagen). Internally labeled cRNA was fragmented by Mg²⁺ hydrolysis in fragmentation buffer (40 mM Tris-acetate, pH 8.2, 100 mM KOAc, 30 mM MgOAc) at 94°C for 35 min. Unlabeled RNA used for end-labeling experiments was transcribed from cDNA using a T7 Megascript kit (Ambion) and fragmented enzymatically or by Mg²⁺ hydrolysis. Because T4 RNA ligase requires a 3'-hydroxyl on the RNA acceptor molecule, fragmented cRNA was dephosphorylated with Shrimp Alkaline Phosphatase (SAP from USB Corp.) prior to ligation. Dephosphorylation also prevents the formation of cRNA concatamers or circularization of the cRNA fragments (19).

Labeling reactions

Labeling reactions were typically performed using 1–10 µg of fragmented RNA, 100–250 µM labeling donor, 90 U T4 RNA ligase (NEB), 16% PEG, 50 mM Tris-HCl, pH 7.8 at 25°C, 10 mM MgCl₂, 10 mM DTT, 1 mM ATP, in a 45 µl volume, and incubated at 37°C for 2 h. Labeling reactions were added directly to hybridization reactions without a cleanup step. ATP was omitted from labeling reactions containing the adenylated labeling reagent AppCpB. The sequence of the RNA oligo model substrate used for ligation optimization is 5'-GUGCC-GUGCCCAGUGGUUCGCAUAA-3'.

Gel shift assay

A streptavidin-biotin gel mobility shift assay was used to indirectly measure ligation efficiency. Following ligation, unincorporated biotin label was removed from the reaction using BioSpin (Biorad) size-exclusion columns. The RNA was then incubated with a molar excess of streptavidin (Pierce) for 10 min before loading on a 4–20, 10 or 20% acrylamide, non-denaturing TBE gel (Invitrogen). The gel was stained

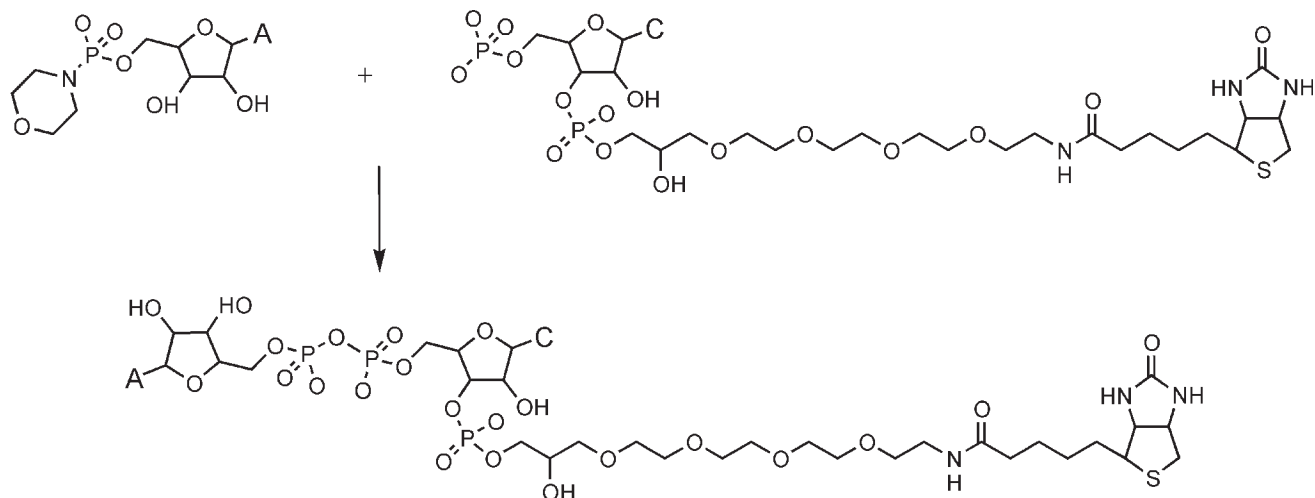


Figure 2. Condensation reaction of adenosine-5'-monophosphoromorpholidate and 5'pCp(TEG-biotin)-3'.

with SybrGold (Molecular Probes) and quantified with AlphaImager software.

Enzymatic fragmentation

Escherichia coli RNase III (a double-strand-specific ribonuclease) was used to fragment total RNA and cRNA. Typically, 10 μ g of RNA were fragmented with 1 U RNase III (New England Biolabs), in 10 mM Tris-HCl, 10 mM MgCl₂, 50 mM NaCl, 1 mM DTT, pH 7.9 at 25°C in a total volume of 20 μ l. Reactions were incubated 35 min at 37°C, followed by 20 min at 65°C. Dephosphorylation was performed concurrently with fragmentation by adding 2 U SAP (USB).

Leukemia RNA samples

Acute lymphoblastic leukemia poly(A) RNA (MOLT-4) was purchased from Clontech. Acute myeloid leukemia total RNA and poly(A) RNA (derived from the KG-1 cell line) were purchased from Ambion. In order to capture the greatest representation of the genes present in the leukemia samples, we performed the IVT amplification protocol originally used to discover the class predictor genes (17,18). Amplified cRNA was generated from AML total RNA (10 μ g) and from ALL poly(A) RNA (5 μ g). For direct labeling experiments 4 μ g of AML and ALL poly(A) RNA and 10 μ g of total AML RNA were fragmented with RNase III and direct-labeled with pCpB₃. pCpB₃ (containing three biotins) was chosen for direct labeling because it produces higher signal intensity and better detection sensitivity than pCpB or pCpB₂. All reactions were done in duplicate. Antisense cRNA targets were hybridized to standard U95Av2 arrays; directly labeled poly(A) and total RNA targets were hybridized to sense versions of the U95Av2 array. The U95Av2 sense array was constructed by reverse complementing the probe sequences of the antisense array. The arrays were hybridized at 50°C and washed according to standard protocols. Arrays were scanned on a GCS 2500 and analyzed with MAS 5.0.

AML class predictor genes were derived previously using HuFl GeneChips (17,18). Published predictor gene accession numbers were matched to probe sets on the U95Av2 array. Twenty-three of the published 25 AML predictor genes that

show elevated expression in AML are queried by 26 probe sets on the U95Av2 array. Three of the genes (L08246, M16038 and W28342) are queried by two probe sets each, which behaved consistently in their absolute calls. Genes that were called absent in one duplicate and present in the other duplicate were conservatively treated as absent.

Microarray methods

Standard (antisense-querying) Human Genome U133A and HG U95Av2 arrays were used for cRNA labeling experiments. Direct labeling of AML and ALL mRNA required synthesis of a sense-querying version of the U95Av2 array. Typically, hybridization reactions contained 10 μ g of labeled RNA incubated at 45 or 50°C for 16 h in 100 mM MES, 1 M NaCl, 20 mM EDTA, 0.01% Tween-20, 0.5 mg/ml BSA, 0.1 mg/ml herring sperm DNA, rotating at 60 r.p.m. The arrays were washed, stained with streptavidin-phycoerythrin conjugate and scanned according to standard Affymetrix protocols (GeneChip[®] Expression Manual, Affymetrix Inc., Santa Clara, CA).

RESULTS

Donor compound synthesis and labeling efficiency

The T4 RNA ligation reaction requires that the donor molecule (label) is of the form 5'-pNp-R-3', where R can be hydrogen, a polynucleotide or label moiety, such as biotin or fluorescein (11,12). In order to maximize labeling efficiency, we tested several different donor molecules (Table 1). Since the rate-limiting step in the ligation reaction appears to be formation of the adenylated donor (20) we synthesized a pre-adenylated donor molecule (AppCpB) in order to accelerate the labeling reaction.

Optimization of the ligation reaction was performed using an RNA oligo (20 nt) and unlabeled cRNA as acceptors. Combinations of ligation reaction temperature (4–42°C), time (0.5–18 h), donor concentration (0.01–1 mM) and T4 RNA ligase concentration (0.1–4 U/ μ l) were tested. We also tested additives purported to improve ligation efficiency (e.g. BSA, DMSO, PEG) (8,21) and found that PEG (15–20%) dramatically improved ligation efficiency. Enzyme and label concentration had large effects on ligation efficiency, however

the activity of the labeled donors also varied as follows: AppCpB > pCpB \geq pApB > pCpB₅ \gg pA₅pB (Table 1). These differences in ligation efficiency are in general agreement with previous reports (20,22).

Under optimal conditions, the RNA oligo was labeled nearly to completion (>95%), as determined by a gel shift assay (Figure 3). However, under the same conditions, only 50–60% of magnesium-hydrolyzed cRNA fragments were directly labeled (Figure 4). Increasing the reaction time,

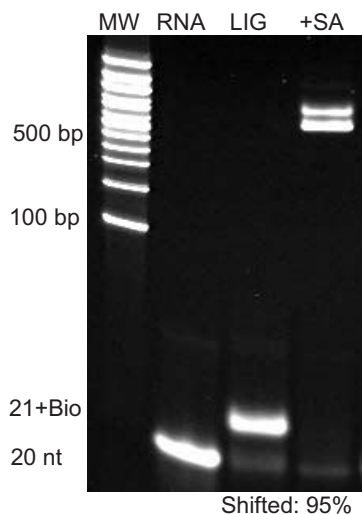


Figure 3. Gel shift analysis of RNA 20mer labeling. The first lane (MW) shows 100 bp Ladder (NEB); the second lane (RNA) contains the 20 nt model RNA substrate before ligation and the third lane (LIG) after ligation to pCpB. Lane +SA contains ligated RNA incubated with streptavidin. Image analysis of lanes 3 and 4 indicate that >95% of the RNA substrate is labeled and shifted. The appearance of two bands in the shifted lane is likely caused by variation in the number of subunits in the streptavidin holoenzyme.

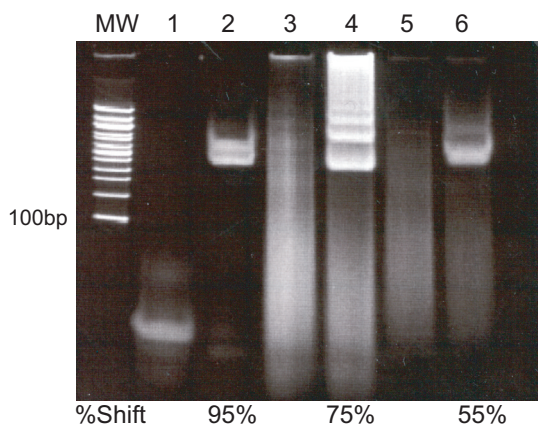


Figure 4. Gel shift analysis of direct-labeled RNA and internally labeled cRNA. RNA 20mer direct-labeled with pCpB (lane 1) and incubated with streptavidin to affect a gel shift (lane 2). Internally labeled cRNA (1 μ g) was fragmented by magnesium hydrolysis (lane 3) and incubated with streptavidin (lane 4). Unlabeled cRNA (500 ng) was fragmented by magnesium hydrolysis and direct-labeled with pCpB (lane 5) and incubated with streptavidin (lane 6). Note that under the same ligation conditions, 95% of the RNA 20mer is labeled and 55% of the magnesium hydrolyzed cRNA is direct labeled. Of the internally labeled cRNA fragments, 75% contain one or more biotin molecule as evidenced by multiple bands in the streptavidin-shifted lane (4).

lowering the reaction temperature or adding additional T4 RNA ligase or SAP did not significantly improve product yield. We developed an unorthodox enzymatic fragmentation method in order to test the effect of fragmentation on ligation efficiency. RNase III, a double-strand-specific endonuclease and S1 nuclease, a single-strand-specific endonuclease, were tested for cRNA digestion. Both enzymes fragmented cRNA, however RNase III produced a fragment size range more similar to the standard method of magnesium hydrolysis (Figure 5). cRNA, total RNA and poly(A) RNA were all fragmented by RNase III and the average fragment size ranged from ~20 to 200 nt. A benefit of RNase fragmentation is that dephosphorylation can be performed simultaneously using SAP. Surprisingly, fragmentation with RNase III dramatically increased the labeling efficiency of cRNA to >90% (Figure 6).

We also fragmented cRNA labeled internally during IVT and tested for the level of biotin incorporation using a gel shift assay (Figure 4). After incubation with streptavidin, 75% of the internally labeled cRNA fragments were shifted indicating they were labeled with biotin. In contrast to directly labeled cRNA, many of the internally labeled fragments appear to contain more than one biotin molecule as evidenced by multiple bands appearing after incubation with streptavidin. This result is expected since the cRNA is labeled by internally

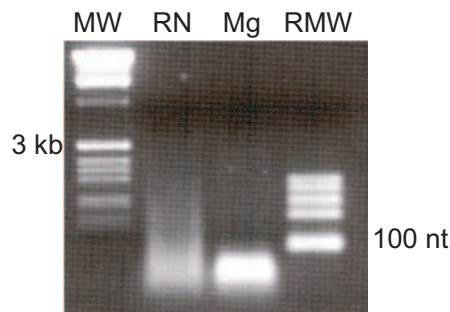


Figure 5. Comparison of cRNA fragmented by RNase III and magnesium hydrolysis. Unlabeled cRNA fragmented with RNase III (RN) has a similar size distribution to cRNA fragmented by magnesium hydrolysis (Mg). Fragments 20–100 nt are ideal for array hybridization.

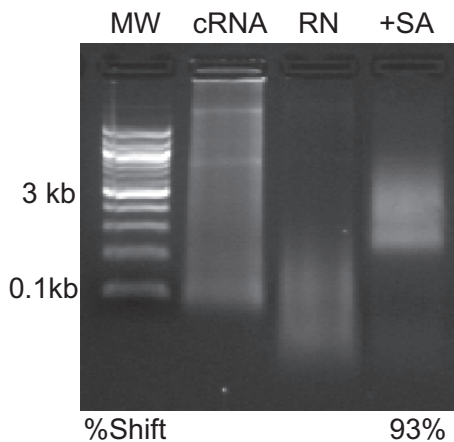


Figure 6. Gel shift assay of cRNA fragmented with RNase III and direct-labeled with pCpB. After RNase III fragmentation and labeling (RN), 93% of the cRNA fragments are shifted when incubated with streptavidin (+SA).

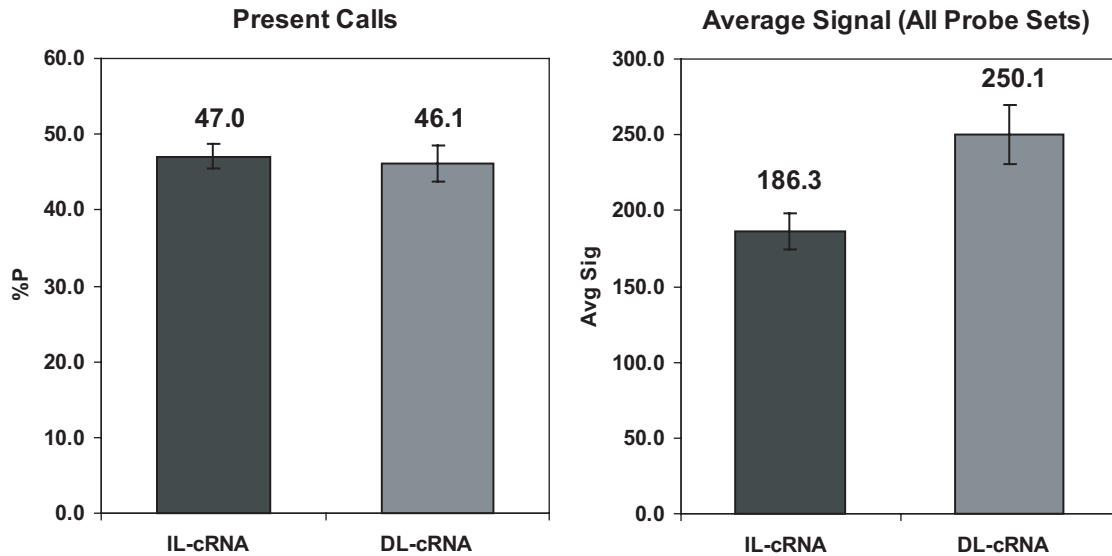


Figure 7. Array performance of direct-labeled (DL) cRNA and internally labeled (IL) cRNA. Triplicate cRNA samples were prepared from total heart RNA and hybridized to U133A arrays under standard conditions (45°C).

incorporating biotin-CTP and biotin-UTP during *in vitro* transcription such that many fragments are likely to contain several biotins.

Array analysis of directly labeled cRNA

Array performance of direct-labeled cRNA was compared to internally labeled cRNA to examine differences between the labeling methods. We used two metrics to gauge array performance; average signal intensity (the mismatch probe intensity subtracted from the perfect-match probe intensity averaged for all probe sets on the array) and absolute present calls (%P, a relative measure of transcript representation and target quality). The algorithm used to derive %P takes several factors into account such as signal intensity, background and hybridization discrimination (23). The average signal intensity of enzymatically fragmented cRNA labeled with pCpB is significantly higher (10–30%) than that of internally labeled cRNA (Figure 7).

Under standard hybridization temperatures (45°C) the %P of internally labeled cRNA is slightly higher than that of direct-labeled cRNA. This effect is likely due to the higher affinity between end-labeled cRNA and the array probes which have been optimized for interaction with internally labeled cRNA. Since internal biotin labels slightly reduce hybridization affinity (6), probe selection algorithms compensate for this interference by designing probes with higher affinity. This high level of affinity results in lower perfect match–mismatch discrimination (used to determine present calls) for end-labeled cRNA. Hybridization of end-labeled cRNA at 50°C significantly improves discrimination and increases present calls. For example, hybridizing the direct-labeled AML poly(A) RNA at 45°C yields 34% present calls whereas hybridization at 50°C yields 37% present calls (Table 2).

Reproducibility for internal labeling and direct labeling methods is very high. Duplicate labeling reactions performed from a common sample of total RNA starting material yield intra-method R^2 correlation coefficients of ≥ 0.98 for both

Table 2. Array performance of directly labeled RNA

Sample	%Present	Total genes present	Average signal
AML IVT-cRNA (anti-sense)	48.5 ± 1.5	6111	210 ± 93.7
AML poly(A) (sense)	37.4 ± 0.4	4712	166 ± 23.4
AML total RNA (sense)	10.6 ± 0.3	1335	69 ± 16.7
ALL IVT-cRNA (anti-sense)	47.1 ± 1.2	5934	244 ± 46.2
ALL poly(A) (sense)	32.1 ± 1.1	4044	124 ± 33.2

methods. An inter-method comparison between directly labeled and internally labeled cRNA yields an R^2 correlation coefficient of 0.94–0.96. A closer examination of call discrepancies (e.g. a probe set is called present in the standard and absent in the directly labeled sample, or vice versa) reveals that 92% of the discordant calls have very low signal intensity (<100) suggesting the transcripts are low in abundance. Further analysis indicates that only 0.63% of the probe sets show a false fold-change ≥ 2 between the labeling methods and only 0.13% show a false fold-change ≥ 3 . In addition, analysis of individual probe sequences indicates that T4 RNA ligase does not exhibit significant labeling sequence bias. These data suggest that the two labeling methods produce highly congruent expression profiles.

Direct labeling with multiple-biotin donor

The labeled donor nucleotides used in this study contain a biotin moiety attached to the 3' phosphate of the donor. Because T4 RNA ligase can utilize donors of varying lengths (14), we hypothesized that extending the tether with multiple biotins would not significantly affect ligation efficiency. pCpB₂, pCpB₃ and pCpB₅ were synthesized containing two, three and five biotins, respectively, attached by TEG linkers. HEG spacer molecules were added between each TEG-biotin to reduce crowding (Table 1, Figure 1). Gel shift analyses indicate that ligation of pCpB₅ is only slightly less efficient than pCpB (one biotin), although more streptavidin is required

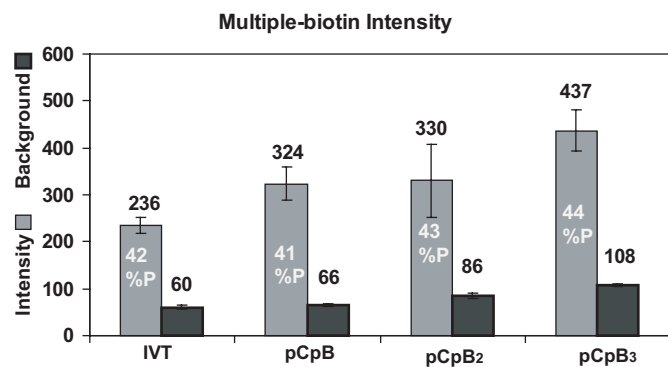


Figure 8. Array performance of cRNA direct-labeled with donor molecules containing multiple biotins. Average signal is increased by labeling with multiple biotins and detection sensitivity (%P) is improved.

to affect a complete gel shift. Donor molecules containing two and three biotin moieties were also efficiently ligated.

Direct labeling of cRNA with donor molecules containing two and three biotins increases average signal intensity and present calls (Figure 8). Labeling target RNA with the five biotin donor molecule resulted in high levels of 'speckling' and anomalous regions of high signal intensity that impaired array performance.

Directly labeling leukemia total RNA and poly(A) RNA

Detection of differentially expressed genes that define a disease state is critical for microarray diagnostics. Since direct labeling does not involve target amplification, detection of low-abundance transcripts is a challenge. Nevertheless, direct labeling might still detect a set of predictor genes sufficient for accurate diagnosis. We tested the ability of our direct-labeling method to detect acute leukemia (AML and ALL) class predictor genes previously discovered by microarray analysis (17,18). AML and ALL poly(A) RNA were directly labeled with pCpB₃ and compared to cRNA created using the IVT amplification method in the original study that derived the class predictor genes. As a more stringent test of sensitivity, AML total RNA was also direct-labeled and profiled. All direct-labeled samples were hybridized to a sense-querying version of the U95Av2 array while cRNA samples were hybridized to standard (anti-sense querying) U95Av2 arrays.

Absolute calls for directly labeled AML poly(A) RNA indicate that 37% of the genes were called present, for a total of 4712 genes detected in the sample (Table 2). Of the genes in the AML cRNA sample, 48% were called present. In the directly labeled ALL poly(A) sample, 32% were called present, for a total of 4044 genes detected, whereas 47% of the genes were called present in ALL cRNA. As expected, directly labeled AML total RNA had very low signal and only 10% of the genes were called present. Nevertheless, this resulted in a total of 1335 genes robustly detected from total RNA. The vast majority of genes (>90%) that were called present in the cRNA samples and absent in the directly labeled poly(A) samples had an average signal of less than 100 fluorescence units, indicating that they are relatively low in abundance.

A comparison of signal intensities was made using the internally labeled AML cRNA as the baseline. Despite an

Table 3. Correlation coefficients of acute leukemia RNA

Comparison	R^2 Correlation ^a	% no change ^b
AML RNA		
cRNA versus cRNA ^c	0.99	99.3
cRNA versus direct-label poly(A) RNA	0.70	86.7
cRNA versus direct-label total RNA	0.05	85.0
ALL RNA		
cRNA versus Std cRNA	0.99	99.6
cRNA versus direct-label poly(A) RNA	0.64	86.2

^aCorrelation of all probe-pair intensity values.

^bPercentage of genes called present in both samples that change less than 2-fold.

^cInternally labeled cRNA.

Table 4. Detection of AML class predictor genes

Accession no. ^a	cRNA	Direct-label poly(A) RNA	Direct-label total RNA
AI362017	A	A	A
L08246	P	P	A
M16038	P	P	P
M23197	P	P	A
M28130	A	A	A
M55150	P	A	A
M62762	P	P	P
M63138	P	P	A
M69043	P	P	A
M80254	P	P	A
M84526	P	A	A
M96326	A	A	A
U46751	P	P	A
U50136	P	P	A
U82759	P	P	P
W28342	A	A	A
X17042	P	P	P
X85116	P	P	P
X95735	P	P	P
Y12670	A	A	A
AL035079	P	P	P
AF020044	P	P	A
X96719	P	P	M
Genes present	20	18	8

^aAccession numbers for predictor genes from the HuFL array used by Golub, *et al.* (17) were matched to the U95Av2 array.

R^2 correlation coefficient of 0.70 between the AML cRNA and the directly labeled poly(A) RNA, 86.7% of the genes called present in both samples showed less than a 2-fold change (Table 3). This translates into 4085 probe sets called 'no change' between the directly labeled poly(A) RNA and the amplified cRNA.

Detection of class predictor genes is more important for diagnostic applications than comprehensive gene detection. We examined genes that generally show elevated expression in AML since these genes should be more easily detected than suppressed genes (presumably low abundance). Using the Affymetrix HuFL array, Golub *et al.* (17) identified a set of class predictor genes that display elevated expression in many, but not all, AML samples tested. Of the 23 class predictor genes contained on the U95Av2 array, 20 were called present in the AML cRNA sample used in this study (Table 4). Since the cRNA amplification protocol efficiently detects low-abundance transcripts (24,25), genes that were not detected

by this method were considered to be absent from our AML sample. Of the 20 remaining AML predictor genes present in the sample, direct labeling of poly(A) AML RNA was able to detect 18, or 90%, of the genes. Examination of the two predictor genes called absent in the poly(A) RNA (M55150 and M84526) indicates that these genes are relatively low in abundance in poly(A) RNA (signal intensities of 58.3 and 89.5, respectively). Eight predictor genes were detected in the directly labeled total RNA sample and the overall signal was significantly lower than in the AML poly(A) RNA sample.

Direct labeling of ALL poly(A) RNA yielded similar results to the AML labeling experiments. Based on previous studies, we analyzed a subset of 32 predictor genes queried on the U95Av2 array that showed elevated expression in many ALL samples (17,18). The IVT amplification protocol called 22 of the genes present. A total of 18 genes were called present in the directly labeled poly(A) RNA, or 82% of the predictor genes considered present in our ALL sample. Of the total number of genes that were called present in both the amplified cRNA and direct-labeled poly(A) RNA, 86.2% showed less than a 2-fold change (Table 3).

DISCUSSION

Performance of directly labeled cRNA

We have developed an efficient system for direct labeling RNA using T4 RNA ligase and 3'-biotinylated nucleotide donors. Directly labeled cRNA displays higher signal intensity and equivalent transcript detection sensitivity compared to cRNA internally labeled during IVT. Uniform labeling, increased target-probe affinity and biotin accessibility likely play a role in the high signal intensities we observe for direct-labeled cRNA.

This direct labeling method can be performed in 2 h in a single reaction tube, making the protocol ideal for automation. The RNA target [cRNA, total RNA or poly(A) RNA] is fragmented with RNase III and simultaneously dephosphorylated. After heat-inactivation and ligation, the reaction is ready for immediate hybridization without purification.

Under optimized ligation conditions, a model oligoribonucleotide substrate is labeled essentially to completion (>95%). However, under the same conditions, only 50–60% of cRNA fragmented by Mg²⁺ hydrolysis is labeled. This level of labeling by T4 RNA ligase is within the range of efficiencies previously reported for various RNA substrates (13). Since cRNA is a mixture of RNA transcripts, this level of labeling may represent the average ligation efficiency of the population of RNA substrates (influenced by sequence and secondary structure). Alternatively, Mg²⁺ hydrolysis may alter the RNA 3'-termini in a way that interferes with subsequent ligation (e.g. depurination).

Fragmentation of cRNA with RNase III results in a dramatic improvement in ligation efficiency (>95% of fragments are labeled). RNase III is a double-strand-specific ribonuclease and it was somewhat unexpected that cRNA was reproducibly fragmented to a size range appropriate for microarray analysis (the majority of fragments are 20–200 nt). Comparison of Mg²⁺-hydrolyzed cRNA and RNase-III-fragmented cRNA indicates that RNase-III-digestion produces a slightly larger fragment size range

and comparable fragment yield (Figure 5). cRNA is a heterogeneous population of RNA, highly enriched in antisense copies of mRNA and evidently contains sufficient stable and transient secondary structure to allow digestion by RNase III. Nuclease S1, a single-strand-specific endonuclease, fragmented cRNA to a lesser extent supporting the idea that a large portion of cRNA is double stranded (data not shown). An additional advantage of RNase-III-fragmentation is that the reaction reaches completion and is less subject to concentration-dependent variability.

The present call rate (%P) of end-labeled cRNA was improved by hybridizing at 50°C, rather than the standard temperature of 45°C. The higher temperature improves discrimination between the perfect match and mismatch probes and increases transcript detection sensitivity (%P). The higher target-probe affinity of directly labeled cRNA suggests that internal biotin labels subtly affect hybridization thermodynamics and duplex stability (6). Designing probe selection algorithms to take advantage of the higher affinity of end-labeled RNA should further improve transcript detection sensitivity.

Labeling with multiple-biotin donors

The direct labeling system we have developed allows attachment of multiple biotins to the 3'-terminus of the nucleotide donor. In theory, each labeled RNA fragment could bind multiple fluorophores, thus increasing signal intensity. We found that labeling cRNA with three biotins increases average signal intensity by up to 30%. This higher signal intensity provided a useful improvement in detection sensitivity (%P), even though there was a concomitant increase in the background (Figure 8). Genes called present with multiple biotin labeling but not with single biotin labeling tend to have relatively low signal (data not shown), suggesting that they are low in abundance. The exact reason for this improvement in sensitivity is still being investigated.

The increase in signal intensity we observe is not proportional to the number of biotins added (i.e. three biotins do not produce a 3-fold increase in average signal). This is likely due to the fact that SAPE stain consists of fluorescent phycoerythrin conjugated to a streptavidin tetramer that can bind more than one biotin (up to four biotins in a fully active tetrameric complex). Nevertheless, the increase in signal intensity observed with the pCpB₃ donor suggests that, on average, fragments can bind more than one SAPE molecule.

cRNA directly labeled with pCpB₅ containing five biotins displayed unusual hybridization patterns and poor array performance. Regions of very high signal intensity that did not conform to the probe feature boundary and a high level of random 'speckling' suggest that target aggregation may have occurred. We are currently testing multiple-biotin donors with different linker configurations and optimizing hybridization buffer composition to alleviate this effect.

Direct labeling of leukemia RNA

Total RNA and poly(A) RNA from acute myeloid leukemia and poly(A) RNA from acute lymphoblastic leukemia cell lines were directly labeled and hybridized to sense versions of the U95Av2 array. Gene expression profiles of directly labeled RNA were examined for the detection of key genes

that were previously found to be robust indicators of leukemia class (17,18). Our rationale for this analysis is that these genes are causative or indicative of the leukemia disease state and therefore some subset should be detectable by direct labeling even though we expect there to be differences in the absolute expression profiles due to the different target preparation methods. We found that direct labeling of poly(A) AML RNA was able to detect 90% of the class predictor genes that were called present using an IVT-based (cRNA) gene expression protocol (see Materials and Methods). Direct labeling of poly(A) ALL RNA was able to detect 84% of the predictor genes called present in the amplified cRNA. As expected, the signal intensity of the AML directly labeled total RNA was significantly lower than the directly labeled poly(A) RNA, reflecting the low abundance of mRNA in the sample. Nevertheless, 1335 genes and 40% of the predictor genes were still called present in the sample.

Comparison of directly labeled AML poly(A) RNA (sense) and AML cRNA (antisense) indicates that the two labeling methods are reasonably concordant. For example, 86.7% of the genes that were called present in both samples showed less than a 2-fold change in abundance (Table 3). The expression profile of directly labeled mRNA probably more accurately represents the initial transcript population because there are fewer sample processing steps (direct labeling requires fragmentation and labeling, compared to at least four enzymatic reactions and two purifications for the IVT amplification protocol). Nevertheless, the high level of concordance between genes called present in both methods suggests that IVT-based mRNA amplification preserves relative transcript abundance.

These results demonstrate the feasibility of using direct labeling of poly(A) RNA, and possibly total RNA, to detect leukemia predictor genes. Despite the fact that the RNA samples used in this proof-of-principle study were derived from leukemia cell lines, we were able to detect a subset of the predictor genes derived from clinical samples. We are currently planning to generate expression profiles for several AML samples to determine if the subset of predictor genes detected by direct labeling is sufficient for reliable tumor classification. Moreover, detailed analysis of a larger set of AML profiles is likely to yield a direct-labeling-specific set of predictor genes that may improve classification robustness.

In addition to diagnostic applications, the utility of direct labeling can be extended to transcriptome analysis. Our labeling method has been successfully used to investigate the role of small non-coding RNAs by directly labeling total RNA and hybridizing the target to whole-genome querying microarrays (26).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Tom Gingeras for valuable comments and Mortezai Vaghefi at Trilink for assistance with synthesis of the donor molecules.

REFERENCES

- Wang, J., Hu, L., Hamilton, S.R., Coombes, K.R. and Zhang, W. (2003) RNA amplification strategies for cDNA microarray experiments. *Biotechniques*, **34**, 394–400.
- Baugh, L.R., Hill, A.A., Brown, E.L. and Hunter, C.P. (2001) Quantitative analysis of mRNA amplification by *in vitro* transcription. *Nucleic Acids Res.*, **29**, e29.
- Gupta, V., Cherkassky, A., Chatis, P., Joseph, R., Johnson, A.L., Broadbent, J., Erickson, T. and DiMeo, J. (2003) Directly labeled mRNA produces highly precise and unbiased differential gene expression data. *Nucleic Acids Res.*, **31**, e13.
- van Belkum, A., Linkels, E., Jelsma, T., van den Berg, F.M. and Quint, W. (1994) Non-isotopic labeling of DNA by newly developed hapten-containing platinum compounds. *Biotechniques*, **16**, 148–153.
- Kelly, J.J., Chernov, B.K., Tovstanovaly, I., Mirzabekov, A.D. and Bavykin, S.G. (2002) Radical-generating coordination complexes as tools for rapid and effective fragmentation and fluorescent labeling of nucleic acids for microchip hybridization. *Anal. Biochem.*, **311**, 103–118.
- Cook, A.F., Vuocolo, E. and Brakel, C.L. (1988) Synthesis and hybridization of a series of biotinylated oligonucleotides. *Nucleic Acids Res.*, **16**, 4077–4095.
- Silber, R., Malathi, V.G. and Hurwitz, J. (1972) Purification and properties of bacteriophage T4-induced RNA ligase. *Proc. Natl Acad. Sci. USA*, **69**, 3009–3013.
- Richardson, R.W. and Gumpert, R.I. (1983) Biotin and fluorescent labeling of RNA using T4 RNA ligase. *Nucleic Acids Res.*, **11**, 6167–6184.
- Hecht, S.M., Alford, B.L., Kuroda, Y. and Kitano, S. (1978) 'Chemical aminoacylation' of tRNA's. *Biol. Chem.*, **253**, 4517–4520.
- Igloi, G.L. (1996) Nonradioactive labeling of RNA. *Anal. Biochem.*, **233**, 124–129.
- Kaufmann, G., Klein, T. and Littauer, U.Z. (1974) T4 RNA ligase: substrate chain length requirements. *FEBS Lett.*, **46**, 271–275.
- Romaniuk, E., McLaughlin, L.W., Neilson, T. and Romaniuk, P.J. (1982) The effect of acceptor oligoribonucleotide sequence on the T4 RNA ligase reaction. *Eur. J. Biochem.*, **125**, 639–643.
- England, T.E., Bruce, A.G. and Uhlenbeck, O.C. (1980) Specific labeling of 3' termini of RNA with T4 RNA ligase. *Methods Enzymol.*, **65**, 65–74.
- England, T.E. and Uhlenbeck, O.C. (1978) Enzymatic oligoribonucleotide synthesis with T4 RNA ligase. *Biochemistry*, **17**, 2069–2076.
- Rosemeyer, V., Laubrock, A. and Seibl, R. (1995) Nonradioactive 3'-end-labeling of RNA molecules of different lengths by terminal deoxynucleotidyltransferase. *Anal. Biochem.*, **224**, 446–449.
- Martin, G. and Keller, W. (1998) Tailing and 3'-end labeling of RNA with yeast poly(A) polymerase and various nucleotides. *RNA*, **4**, 226–230.
- Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A. et al. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531–537.
- Armstrong, S.A., Staunton, J.E., Silverman, L.B., Pieters, R., den Boer, M.L., Minden, M.D., Sallan, S.E., Lander, E.S., Golub, T.R. and Korsmeyer, S.J. (2002) MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nature Genet.*, **30**, 41–47.
- Uhlenbeck, O.C. and Cameron, V. (1977) Equimolar addition of oligoribonucleotides with T4 RNA ligase. *Nucleic Acids Res.*, **4**, 85–98.
- Hoffmann, P.U. and McLaughlin, L.W. (1987) Synthesis and reactivity of intermediates formed in the T4 RNA ligase reaction. *Nucleic Acids Res.*, **15**, 5289–5303.
- Tessier, D.C., Brousseau, R. and Vernet, T. (1986) Ligation of single-stranded oligodeoxyribonucleotides by T4 RNA ligase. *Anal. Biochem.*, **158**, 171–178.
- McLaughlin, L.W., Piel, N. and Graeser, E. (1985) Donor activation in the T4 RNA ligase reaction. *Biochemistry*, **24**, 267–273.
- Liu, W.-m., Mei, R., Di, X., Ryder, T.B., Hubbell, E., Dee, S., Webster, T.A., Harrington, C.A., Ho, M.-h., Baid, J. and Smeekens, S.P. (2002) Analysis of high density expression microarrays with signed-rank call algorithms. *Bioinformatics*, **18**, 1593–1599.
- Ghasemzadeh, M.B., Sharma, S., Surmeier, D.J., Eberwine, J.H. and Chesselet, M.F. (1996) Multiplicity of glutamate receptor subunits in single striatal neurons: an RNA amplification study. *Mol. Pharmacol.*, **49**, 852–859.

25. Lockhart,D.J., Dong,H., Byrne,M.C., Follettie,M.T., Gallo,M.V., Chee,M.S., Mittmann,M., Wang,C., Kobayashi,M., Horton,H. *et al.* (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.*, **14**, 1675–1680.
26. Kampa,D., Cheng,J., Kapranov,P., Yamanaka,M., Brubaker,S., Cawley,S., Drenkow,J., Piccolboni,A., Bekiranov,S., Helt,G. *et al.* (2004) Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.*, **14**, 331–342.