Hybrid selection with cDNA–silica

Harry W. Jarrett

Department of Biochemistry, 858 Madison Avenue, University of Tennessee, Memphis, TN 38163, USA

Abstract

A partial length ovalbumin cDNA–silica was produced using primer extension of (dT)18–silica with annealed partial ovalbumin RNA and reverse transcriptase. This cDNA–silica was used to test whether full-length ovalbumin RNA could be selectively purified in the presence of a large excess of other (mouse muscle) RNA. The cDNA–silica synthesized had minimally 60 pmol cDNA per gram silica and had a capacity for full-length ovalbumin RNA of minimally 38 μg/g. Even when other RNA was present in greater than 1000-fold excess, ovalbumin RNA was selectively retained by the cDNA–silica and was eluted in yields of 43% with an enrichment which varied over the range of 29–162-fold in various experiments. These results show that even rare RNAs can be selectively purified in high yield using cDNA–silica. The importance of these results to hybrid selection and subtractive library preparation is discussed.

Keywords: Hybrid selection; cDNA–silica; Affinity adsorbents

1. Introduction

In our previous report [1], we have shown that RNA serves as an efficient template for the enzymatic production of cDNA–silica. In this procedure, silica is chemically coupled to the 5' end of 5'-aminoethyl–(dT)$_{18}$. The oligothymidylic acid is then annealed to 3' polyadenylated RNA (poly A–RNA) which acts as a template for production of a cDNA strand by extension of the (dT)$_{18}$ "primer". The RNA template is then eluted to produce a single stranded cDNA–silica. The resulting cDNA was shown to be primarily full length and of the correct sequence by direct solid phase sequencing [1].

DNA–silicas have been shown to anneal with complementary strands and hybridization follows similar temperature-dependent dissociation ("temperature of melting", T_m) as does hybridization in the liquid phase [2]. Elution by decreasing salt concentration, increasing temperature, or denaturation with an organic solvent mobile phase are all capable of highly selective elution, resolving oligonucleotides differing by a single nucleotide in length [2,3].

DNA-supports have been used for the selective affinity chromatography of transcription factors and other DNA-binding proteins ([4]; for review see Refs. [5,6]), for selection of mRNA by poly A tail length [3,7,8], and for hybrid selection of specific DNA and RNA sequences [1,2,9–13]. One interesting application of hybrid selection is the preparation of subtractive libraries [10–13]. In a typical variation of this technique, cDNA produced from one cell type is immobilized on a solid support and hybridized with the mRNA of another cell. The RNA which does not hybridize is recovered and is consequently

enriched in sequences unique to the second cell type. This RNA can then be used to construct a library of cell-specific genes [10–13].

If cell-specific subtractive libraries are to be efficiently constructed, several conditions must pertain. The immobilized cDNA must efficiently bind all of the corresponding mRNA common to both cell types. Any "leakage" results in a mixed library containing sequences derived from both cell types and selection of cell specific genes becomes more tedious. Furthermore, sequences can compete with one another for DNA-support binding. If more abundant mRNA sequences compete with rare ones, a library enriched in less abundant transcripts could result. Since such transcripts should be weighted in favor of rare transcripts common to both cell lines and detection of rare sequences can be challenging, this potential bias in library construction could be especially misleading.

Another potential use of hybrid selection is to use a specific cDNA to purify its complementary RNA. This could be especially useful when a partial cDNA is obtained by library screening. The cDNA can be immobilized, used to select for the corresponding full-length mRNA and this RNA used in in vitro translation with other techniques to demonstrate the authenticity of the partial cDNA. Furthermore, the full length mRNA could be cloned to construct a library enriched in the full-length cDNA of interest.

Since the enzymatic synthesis of cDNA-supports such as cDNA–silica [1], cDNA–latex [10–12], and cDNA–paramagnetic beads [13] has utilized oligothymidine primers and poly A–RNA templates, these supports can potentially bind any poly A–mRNA by way of hybridization of the poly A sequences with the support oligothymidylic acid. These hybrids should have a low melting temperature and could presumably be removed by washing the support under stringent conditions. The latter two supports have been used for subtractive library construction [10–13]. However, the depletion of sequences caused by poly A hybridization, the forms of "leakage" discussed above, or rare sequence enrichment has not so far been assessed.

Here, we present simple experiments which measure the contribution of some of these factors to successful hybrid selection. These experiments utilize HPLC-grade cDNA–silica which has excellent

mass transfer characteristics, however, the experiments and the results obtained are probably applicable to other supports. The results suggest ways of improving the selectivity for cell specific genes and are the first to quantify recovery of rare sequences.

2. Methods

2.1. Ovalbumin RNA synthesis

The partial ovalbumin RNA and its use as template for cDNA–silica synthesis was described previously [1]. The resulting cDNA–silicas were sequenced as previously described [1]. This partial RNA contains 340 bp of the full-length ovalbumin's 1.9 kb RNA, containing primarily sequences from the 5' end of the RNA with a large internal deletion, followed by a short sequence from the 3' end and a stretch of about 30 adenylate residues (i.e., poly A tail). For some experiments, the synthesis was carried out using α-^{35}S-rUTP to yield a labeled RNA template to follow template elution during cDNA–silica synthesis.

Full-length ovalbumin RNA was also prepared in the same way (i.e., using reagents and protocols suggested by Promega as modified in Ref. [1]) except using the full-length ovalbumin cDNA in pGEM3Zf(−), relaxed by HindIII restriction, as template. Synthesis was usually carried out using α-^{35}S-rUTP and the labeled, full-length RNA was used to test binding and elution to the various partial ovalbumin cDNA–silicas.

2.2. Mouse RNA

Mouse muscle RNA was isolated using the guanidinium thiocyanate method as previously described [14] and labeled by the method of Cobianchi and Wilson [15] using 5'-^{32}P-3',5'-cytidine bisphosphate (^{32}P-pCp) and T_4 RNA ligase.

2.3. Determination of $(dT)_{18}$ content of silicas

To determine the $(dT)_{18}$ content of silica, it was annealed to an excess amount of lac oliconucleotide

(GTGGAATTGTGAGCGGATAACAATT-(dA)$_{18}$)
which had been 5' end labeled with γ-^{32}P-ATP and
T$_4$ polynucleotide kinase. After thorough washing,
the oligonucleotide was eluted (42°C in TE buffer)
and scintillation counted. As previously reported
[16], this method gives results which agree well with
direct determination of DNA content by hydrolysis
and phosphate assay.

2.4. Determination of the amount of full length cDNA-silica

After reverse transcription to produce the cDNA–
silica, it was probed with excess 5' end labeled
OV–S oligonucleotide (GCTAGAAAGCUG-
TATTGCC). This oligonucleotide specifically hy-
bridizes with the sequence between 8–26 nucleotides
from the 3' end of the ovalbumin cDNA and detects
the amount of cDNA which has the complete,
template directed sequence.

3. Results

The bulk synthesis of (dT)$_{18}$-silica has been
described [9]. For this study, this synthesis was
carried out on four different occasions and yielded
silicas containing 2.2, 5.7, 12.3, and 16.6 nmol
(dT)$_{18}$/g silica. All four were successfully used in
the course of the experiments presented here for
producing cDNA silica.

A typical cDNA–silica synthesis is presented in
Fig. 1 except that in this case, the template RNA
strand was synthesized with ^{35}S-UTP so that its
binding to silica and elution after synthesis could be
easily followed. A large excess of the template
strand is added, heated and cooled to allow anneal-
ing, and thoroughly washed. Reverse transcriptase is
then added and synthesis occurs. If synthesis was
unsuccessful, the template elutes in 42°C TE (data
not shown but see below); if successful, high tem-
perature washes (90°C TE) are required to elute the
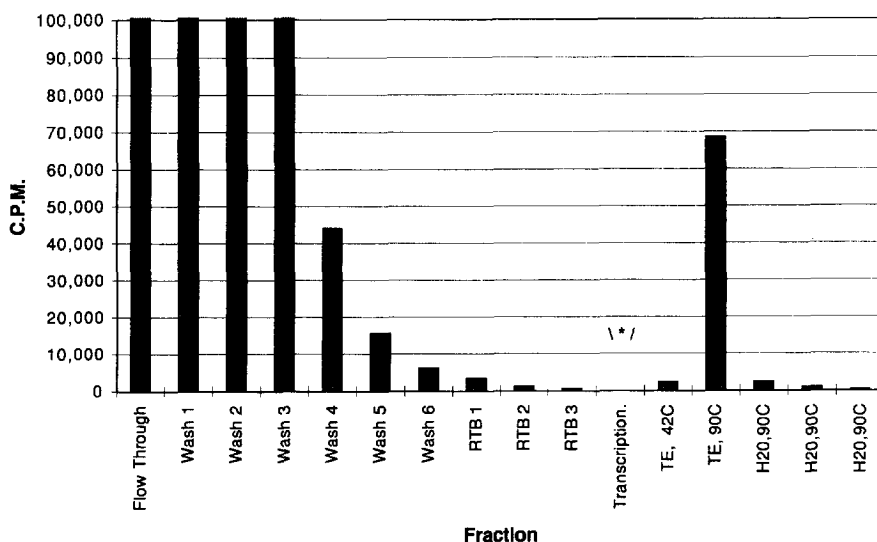template as occurs in Fig. 1.

Fig. 1. Synthesis of partial ovalbumin cDNA–silica. To 50 mg (dT)$_{18}$–silica (12.3 nmol (dT)$_{18}$/g. silica) was added 1 nmol of ^{35}S-rUTP labeled partial ovalbumin RNA. The silica was heated to 65°C and slowly cooled to room temperature over 30 min. The silica was then pelletted, the supernatant ("flow through") mixed with scintillation fluid and counted. The silica was then washed six times with 0.2 ml reverse transcriptase (RT) buffer (50 mM Tris–HCl, pH 8.3, 75 mM KCl, 3 mM MgCl$_2$, 10 mM dithiothreitol) and then three more times with 0.2 ml of RT buffer containing 2 mM of each dNTP. The supernatant from the last wash was removed and 12 μl of Moloney murine leukemia virus reverse transcriptase (2400 units) was added and the mixture incubated at 37°C for 1 h (indicated by "*" on figure). For subsequent washes, each 0.2-ml wash was repeated three times and the pooled 0.6 ml was counted. The silica was then washed at 42°C with TE (10 mM Tris–HCl, pH 7.5, 1 mM EDTA), 90°C TE, and finally with 90°C H$_2$O.

This synthesis was carried out six times for these studies. For two of these syntheses, the amount of ovalbumin RNA which binds and the amount of full-length cDNA produced were both determined. Although less complete data was obtained for the other silicas, the results in Table 1 are representative of what was found in other syntheses. For most syntheses, 0.1–0.3 nmol of cDNA were produced per gram of silica by enzymatic synthesis. The cause of the variability is not known but may result from variations in the quality of the RNA template synthesized and differences in the efficiency of mixing during the synthesis reaction. The cDNA–silicas were also sequenced as previously described [1] and have the expected sequence. The results in Table 1 point out some important characteristics of these silicas. Only a fraction of the $(dT)_{18}$ present in the initial silica is capable of binding poly A RNA and extension by reverse transcriptase. The reason for this is unclear but may result from steric crowding. These porous silicas have 30-nm pores and surface areas of ~100 m^2/g. Considering only the dimensions of double stranded polynucleotides, it would be possible to pack ~40 μmol/g into this area but less than 1:10,000 of this amount is observed in our experiments. Once extension does occur, at least a third and usually more is extended to full length. The amounts of full-length are adequate for most types of experiments. For example, the lower amount of full-length (Table 1, experiment 2, 0.06 nmol/g.)

Table 1
Only a fraction of $(dT)_{18}$ primes cDNA–silica synthesis but synthesis is primarily full-length

Experiment	$(dT)_{18}$ content[a]	cDNA-content[b]	Full length[c]
1	16.6±1	0.36±0.06	0.13±0.02
2	12.3±1.1	0.080±0.04	0.060±0.001

[a] All values given are in nmol nucleic acid per gram silica.The initial amount of $(dT)_{18}$ chemically coupled to silica determined by hybridization with the lac–$(dA)_{18}$.

[b] After the reverse transcriptase reaction and removal of the template RNA strand, excess ^{35}S-labeled ovalbumin RNA (partial) was annealed. After thorough washing, the labeled RNA was eluted and quantified by scintillation counting.

[c] The single stranded cDNA–silica was also probed with an excess of the OV–S oligonucleotide which specifically binds to the 3' end of the cDNA and thus detects the full length cDNA product. The amount eluting after thorough washing is shown.
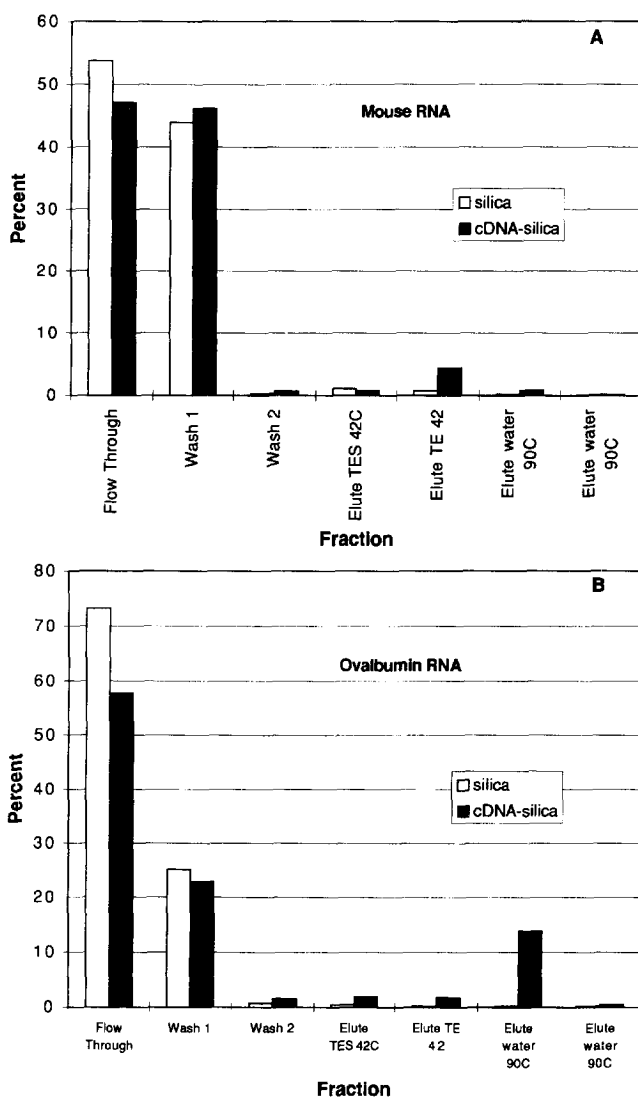
is equivalent to 38 μg of the 1.9 kb ovalbumin RNA per gram silica.

Since cDNA–silica synthesis relies upon poly A template binding, essentially any poly A sequence would be expected to bind and we were curious to know if this affects the usefulness of this material for binding selectively a specific RNA of interest. The experiment in Fig. 2 addresses this point. When mouse RNA is annealed to $(dT)_{18}$–silica or to the partial ovalbumin cDNA–silica, only a fraction binds. This bound RNA elutes at 42°C in the TE buffer. The short hybrid between $(dT)_{18}$ and poly A RNA would be expected to dissociate at this low temperature. That effective elution of the cDNA silica also required lowering the salt concentration (i.e., TES to TE) suggests that the reverse transcriptase reaction lengthened the (dT) sequence on the silica using the ~30 nt. poly A tail on the ovalbumin RNA template.

The result was quite different when the full-length ovalbumin RNA was applied to these silicas. In this case, the cDNA–silica was only eluted when the temperature was increased to 90°C. The partial cDNA on the silica would be capable of forming a 340 bp. hybrid with the applied, full-length ovalbumin RNA and this hybrid only dissociates at high temperatures and low salt.

Fig. 3 shows an experiment to test the selectivity of, the partial ovalbumin cDNA–silica. Full-length ovalbumin mRNA was synthesized using high specific activity α-^{35}S-rUTP and mixed with a 1133-fold excess of $^{32}PO_4$-pCp labeled mouse RNA. Approximately the same number of counts of each isotope was present in the sample to allow accurate determination of both isotopes by scintillation counting. This sample was applied to $(dT)_{18}$–silica treated identically to the ovalbumin cDNA–silica throughout except that no reverse transcriptase was added during its synthesis and thus no template directed cDNA sequences were produced. The result was that a fraction of both the ovalbumin and mouse RNA are retained by the column and elute together in the 150 mM NaCl containing buffer at 42°C. This is as expected since both ovalbumin and mouse mRNA have 3' poly A tail sequences capable of hybrid formation with the $(dT)_{18}$ sequences on the support and this melting temperature would be expected for short hybrids at this salt concentration. Furthermore,

Fig. 2. The elution behavior of full-length ovalbumin RNA differs from other (mouse) RNA. Approximately equal counts of ^{32}P-pCp labeled mouse RNA (1.3×10^5 c.p.m., 5.6 μg, panel A) or ^{35}S-UTP labeled, full-length ovalbumin RNA (1.5×10^5 c.p.m., 0.04 μg, panel B) in TES buffer (10 mM Tris–HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA) was annealed (see Fig. 1) to 5 mg of either partial ovalbumin cDNA–silica ("cDNA–silica", solid bars) or silica prepared in the same way except that reverse transcriptase was not added ("silica", open bars). The silica was then thoroughly washed with TES buffer at room temperature (19°C), with TES at 42°C, with TE at 42°C, and finally with water at 90°C. As in Fig. 1, each 0.2-ml wash was repeated three times and pooled to give the 0.6 ml which was counted and shown in the figure.

the ratio of ovalbumin/mouse RNA in the eluted fraction (1:307) was not markedly different from that present in the applied sample (initially 1:1133). No selection was expected since poly A sequences for the two RNAs should not be appreciably different. However, when the same mixture was applied to the

partial ovalbumin cDNA–silica, the result was markedly different. In this case, while most of the mouse mRNA passes through the column unretained, most of the ovalbumin mRNA is retained and is subsequently eluted. Furthermore, the conditions under which it elutes are different from the previous
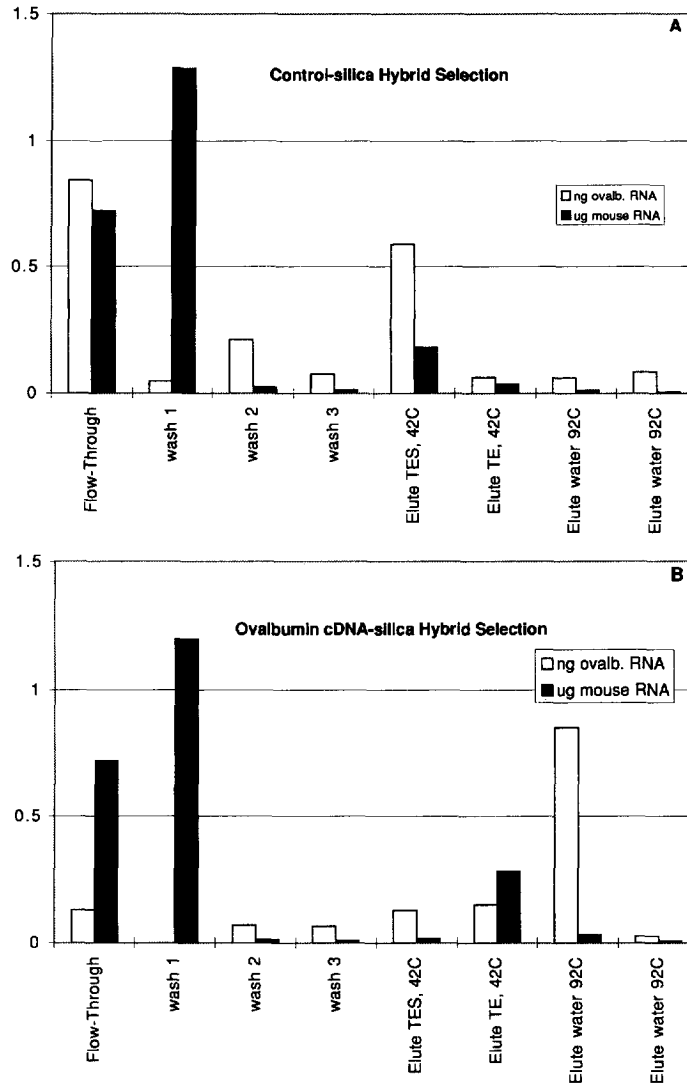
Fig. 3. Even rare RNAs are efficiently selected by cDNA–silica. 2.3 μg of ^{32}P-pCp labeled mouse muscle RNA (7×10^5 c.p.m., closed bars) was mixed with 2 ng of full-length ^{35}S-labeled ovalbumin RNA (1.4×10^5 c.p.m., open bars) in TES buffer and annealed (see Fig. 1) to 100 mg of the partial ovalbumin cDNA–silica (panel B) or to silica treated identically except that reverse transcriptase was not added for synthesis (panel A). After thorough washing in TES buffer at room temperature, the silica was selectively eluted with TES at 42°C, with TE at 42°C, and finally with water at 92°C. As in the previous figures, each 0.6-ml wash was the pool of three 0.2-ml washes. The fractions were counted on a dual channel scintillation counter and the amounts of ^{32}P-mouse RNA and ^{35}S-ovalbumin RNA were calculated and are shown in the figure.

experiment. While most of the mouse RNA retained by the silica elutes at 42°C, elution of the ovalbumin RNA occurs only when 92°C water is used. The eluted ovalbumin RNA is also markedly enriched over the starting material. While present initially as

only ~0.1% of the total RNA, it now comprises nearly 3%, an enrichment of 29-fold. Further elution of the silica under these conditions elutes little additional RNA. This experiment was carried out several times and enrichment as high as 162-fold was

obtained in experiments using 5 mg silica and the same washing volumes as Fig. 3.

4. Discussion

The results obtained have important implications for the various uses of hybrid selection. Firstly, RNAs present in only low abundance in a sample are retained by DNA–silica even in the presence of a large excess of other RNA sequences (Fig. 3). Competition between different RNAs would apparently not select against these less abundant RNAs. This probably also hold true when cDNA–silica is prepared with a heterogeneous mRNA template as would be the case in subtractive library construction, however, the kinetics of such an experiment should also be considered. When two polynucleotides hybridize, the second-order rate of association depends upon the concentration of both strands as well as their complexity [17]. When rare sequences present in one cell type are used for cDNA–support construction and used to select common sequences from a second cell type, hybridization of rare sequences will be slower than more abundant ones and additional hybridization time will be required. Here, hybridization occurred relatively quickly ($\leq$30 min) when cDNA–silica sequences were present at high concentration. When less cDNA is present on the support, either because of inefficient synthesis or because of natural abundance in a heterogeneous template, hybridization would occur more slowly and longer times would be required if the resulting library is to be representative of the cell types.

However, selection for a given sequence is not great. Purification in the range of 29–162-fold was obtained here and several experiments using stringent washing did not improve this by more than a factor of about 2-fold (data not shown). The reason greater purity was not obtained is unclear. It is unlikely that skeletal muscle produces ovalbumin RNA, certainly not enough to have had a biasing effect on the experiment. The support which is the starting material for $(dT)_{18}$–silica [2] is a succinic acid based weak cation exchanger [18], now called WCX–silica (Alltech Associates), and should actually repel RNA at low salt concentrations and so secondary chromatographic modes operating to re-

tain RNA are unlikely. The wash and elution steps shown in Fig. 3 each represents three different washes of 0.2 ml which were combined. If mass transfer of unretained RNA were complete, each combined wash should remove minimally about 88% of the unretained material at each of the washes shown. Washing prior to elution of the ovalbumin RNA should have resulted in retention of less than 3×10^{-5} of the mouse RNA or about 0.3 ng and not the 50 ng actually found in the 92°C eluate. Thus, the washing procedure was sufficient to ensure high purity. These arguments suggest that mass transfer was not complete under the conditions used. Macroporous silica has better performance in this regard than soft gels and other supports which have previously been used for hybrid selection and these other materials would be expected to show lower or equal selectivity.

However, repetition of selection could be used to improve purity. By reapplying samples for a second round of selection with fresh cDNA–silica or after the silica has been stripped with denaturants would allow greater purification. Three repetitions should yield a purification of 24 000-fold (29^3). Initial experiments have demonstrated that greater purification is in fact obtained when selection is repeated with fresh silica (data not shown). Since even rare RNAs are retained by selection, repetition should yield an RNA without altering the abundance of any particular RNA.

The retention of rare sequences in good yield is important for experiments in which a partial cDNA is used to enrich a rare, full-length RNA. This kind of hybrid selection can give recoveries as high as 40% or more even for transcripts present as only 0.1% of total RNA. This may be the most powerful use for cDNA–silica.

Whether cDNA–silica or any similar material can be used to isolate cell-specific mRNA however is questionable. The reason for this involves what are usually called ectopic transcripts. It has been found that even tissues which are not usually thought to express a certain protein do in fact express a small amount of the appropriate mRNA which is indistinguishable in most cases from the authentic transcript obtained from highly expressing cells ([19,20]; reviewed in Ref. [21]). The presence of these rare ectopic transcripts and the ability of cDNA–silica to

selectively retain even rare transcripts suggest that hybrid selection for preparation of subtractive libraries may not be capable of yielding "cell-specific" libraries because, in fact, transcripts themselves are not completely cell-specific.

That subtractive libraries are in fact useful relies rather on a phenomenon related to capacity. If the cDNA (derived from a first cell type) and mRNA (from a second cell type) are used in roughly equal amounts, any more abundant mRNA present (in the second cell type) will be recovered because it exceeds the capacity of the cDNA for further hybridization. However, it is the practice in most such experiments to use an excess of cDNA probe and repetition. In such cases, selection would be for those RNAs present in very high abundance in the second cell type and would be representative of only the gross differences between the two types of cells.

Acknowledgments

References

[1] H.W. Jarrett, J. Chromatogr. A, 708 (1995) 13–18.
[2] T.A. Goss, M. Bard and H.W. Jarrett, J. Chromatogr., 508 (1990) 279–287.
[3] T.A. Goss, M. Bard and H.W. Jarrett, J. Chromatogr., 588 (1991) 157–164.
[4] C. DiRusso, R.P. Rogers and H.W. Jarrett, J. Chromatogr., 677 (1994) 45–52.
[5] H.W. Jarrett, J. Chromatogr., 618 (1993) 315–339.
[6] J.T. Kadonaga, Methods Enzymol., 208 (1991) 10–23.
[7] A.J. Jacobson, R.A. Firtel and H.F. Lodish, Proc. Natl. Acad. Sci. U.S.A., 71 (1974) 1607–1611.
[8] S.L. Phillips, C. Tse, I. Serventi and N. Hynes, J. Bacteriol., 138 (1979) 542–551.
[9] L.R. Solomon, L.R. Massom and H.W. Jarrett, Anal. Biochem., 203 (1992) 58–69.
[10] E. Hara, T. Kato, S. Nakada, S. Sekiya and K. Oda, Nucleic Acids Res., 19 (1991) 7097–7104.
[11] E. Hara, T. Yamaguchi, H. Tahara, N. Tsuyama, H. Tsurui, T. Ide and K. Oda, Anal. Biochem., 214 (1993) 58–64.
[12] K. Kuribayashi-Ohta, S. Tamatsukuri, M. Hikata, C. Miyamoto and Y. Furuichi, Biochim. Biophys. Acta, 1156 (1993) 204–212.
[13] I.R. Rodriguez and G.J. Chader, Nucleic Acids Res., 20 (1992) 3528.
[14] H.W. Jarrett and J.L. Foster, J. Biol. Chem., 270 (1995) 5578–5586.
[15] F. Cobianchi and S. Wilson, Methods Enzymol., 152 (1987) 103–110.
[16] L.R. Massom and H.W. Jarrett, J. Chromatogr., 600 (1992) 221–228.
[17] R.B. Wallace and C.G. Miyada, Methods Enzymol., 152 (1987) 432–442.
[18] H.W. Jarrett, J. Chromatogr., 405 (1987) 179–189.
[19] J. Chelly, J.C. Kaplan, P. Maire, S. Gautron and A. Kahn, Nature, 333 (1988) 858–860.
[20] G. Sarkar and S.S. Sommer, Science, 244 (1989) 331–334.
[21] D.N. Cooper and M. Krawczak, in: Human Gene Mutation, BIOS Scientific Publishers, Oxford, 1993, pp. 61–84,.