Transcript Abundance in Yeast Varies over Six Orders of Magnitude*

Received for publication, February 15, 2002, and in revised form, March 4, 2002
Published, JBC Papers in Press, March 6, 2002, DOI 10.1074/jbc.C200101200

Michael J. Holland‡
From the Department of Biological Chemistry, School of Medicine, University of California, Davis, California 95616

In the current era of functional genomics, it is remarkable that the intracellular range of transcript abundance is largely unknown. For the yeast Saccharomyces cerevisiae, hybridization-based complexity analysis and SAGE analysis showed that the majority of yeast mRNAs are present at one or fewer copies per cell; however, neither method provides an accurate estimate of the full range of low abundance transcripts. Here we examine the range of intracellular transcript abundance in yeast using kinetically monitored, reverse transcriptase-initiated PCR (kRT-PCR). Steady-state transcript levels encoded by all 65 genes on the left arm of chromosome III and 185 transcription factor genes are quantitated. Abundant transcripts encoded by glycolytic genes, previously quantitated by kRT-PCR, are present at a few hundred copies per cell whereas genes encoding physiologically important transcription factors are expressed at levels as low as one-thousandth transcript per cell. Of the genes assessed, only the silent mating type loci, HML and HMR, are transcriptionally silent. The results show that transcript abundance in yeast varies over six orders of magnitude. Finally, kRT-PCR, cDNA microarray, and high density oligonucleotide array assays are compared for their ability to detect and quantify the complete yeast transcriptome.

Measurements of the intracellular range of transcript abundance relied initially on hybridization-based complexity analysis and more recently on SAGE1 analysis. For the yeast Saccharomyces cerevisiae, hybridization-based complexity analysis (1) and SAGE analysis (2) showed that 75% of poly(A) mRNA is encoded by only 20% of yeast genes. SAGE analysis also showed that 75% of yeast genes are expressed at 1 or fewer copies per cell (2). Because of technical limitations, neither method provides an accurate estimate of the range of low abundance transcripts encoded by the majority of yeast genes. We previously demonstrated the accuracy, sensitivity, and reliability of kRT-PCR for quantitating mRNA levels in complex mixtures of total cellular RNA over a wide range of relative transcript abundance (3–5). In contrast to second order hybridization-based complexity analysis or SAGE analysis, where signal to noise decreases exponentially with decreasing transcript abundance, signal to noise is constant for kRT-PCR; only the PCR cycle number where product accumulation is detected varies with transcript abundance. For highly expressed yeast metabolic genes, mRNA levels determined by kRT-PCR are in good agreement (within 2-fold) with those made by Northern blotting, enzyme activity measurements, and SAGE (5). For highly repressed genes, fold repression measured by kRT-PCR versus enzyme activity are within 2-fold down to transcript levels of 0.01 copy per cell (5). Here we employ kRT-PCR to assess the full range of transcript abundance in yeast using selected subsets of the yeast transcriptome and total cellular RNA isolated from early log phase cultures of strain BY4742 (derived from strain S288C) grown in YPD medium.

EXPERIMENTAL PROCEDURES

Each kRT-PCR assay was performed in a 20-μl reaction tube containing: 50 mM Tricine buffer, pH 8.3, 110 mM potassium acetate, 13% glycerol, 0.3 mM dATP, dCTP, and dGTP, 0.05 mM dTTP, 0.5 mM dUTP, 2.4 mM Mn(OAc)2, 2.5 μM ethidium bromide, 0.25 μM primers, 4 units of Tth DNA polymerase, 2 units of uracil N-glycosylase, and 120 ng of rTh DNA polymerase, 2 units of uracil N-glycosylase, and 120 ng of total yeast cellular RNA template. Yeast strains were grown to early log phase in YPD medium containing 2% glucose. Total cellular RNA was extracted from glass bead-disrupted cells and treated with RNase-free DNase I to eliminate residual genomic DNA (5). Primer pair design, instrumentation, and data analysis were as described previously (5). Transcript copy numbers per cell reported here are the average of 15 independent kRT-PCR assays for each mRNA.

RESULTS

To assess the range of expression levels for genes along a single chromosome, transcripts from all 65 genes and computationally annotated ORFs on the left arm of yeast chromosome III were quantitated. Steady-state transcript copy number varied by more than four orders of magnitude, from GLK1 (glucokinase) at 17 copies per cell to YCL069 at 0.001 copy per cell (Fig. 1). With the exception of low abundance transcripts encoded by six ORFs located between the telomere and the HML locus (0.04–0.001 transcript per cell), no obvious clustering of genes encoding high versus low abundance transcripts was detected. The majority of abundant transcripts (2.2–17 transcripts per cell) encoded by 12 genes on the left arm of chromosome III code for metabolic enzymes.

In contrast to mRNAs encoding metabolic enzymes, those encoding transcription factors should be representative of low abundance transcripts. As expected, steady-state levels for 185 transcription factor genes varied over four orders of magnitude with the majority (82%) present at 1 or less transcript per yeast cell (Fig. 2). GCN4 encodes a leucine zipper transcription factor involved in general amino acid control and was the most abundant at 7 copies per cell. The relatively high level of GCN4 mRNA may compensate for the fact that translation of this mRNA in vivo is under strong negative control (6). SPT15, which encodes TBP (TATA-binding protein), required for all nuclear transcription (7) was present at 0.8 copy per cell.
whereas the SUA7 transcript, encoding TFIIB, required for all transcription by RNA polymerase II (8) was present at 4 copies per cell. Transcripts encoding TAFs (TBP associated factors) (9), subunits of the SRB-mediator complex (9) and the swi-snf complex (9) ranged from 2 to 0.1 copies per cell with the majority present at 1 to 0.5 transcripts per cell. Transcripts encoding eight members of the basic helix-loop-helix family of transcription factors (TYE7, INO4, RTG3, PHO4, INO2, RTG1, HMS1, and CBFI) ranged from 3 to 0.09 transcripts/cell. IXR1, GAT1, GLN3, SWI6, and SPT3 (0.0009, 0.002, 0.004, 0.005, and 0.005 transcripts per cell, respectively) encoded very low transcript levels.

To fully assess low level transcription in yeast cells, expression of the mating-type specific regulatory genes was determined using total cellular RNA isolated from haploid and diploid cells (Fig. 3). Expression of the regulatory genes present at the silent mating type cassette HMLo measured in mating type a haploid cells or the silent mating type cassette HMRu measured in mating type a haploid cells was not detected using the kRT-PCR assay (less than 0.0001 transcript per cell). Similarly, a1 expression in diploid cells was not detected using the kRT-PCR assay. Expression of the regulatory genes at the MAT locus was very similar for the strains derived from S288C or D27310B. Unexpectedly, expression of a1 and a2 in diploid cells was 4- to 5-fold lower than in mating type a haploid cells or mating type a haploid cells, respectively.

Hybridization-based cDNA microarray and high density oligonucleotide array technologies are widely used for transcript profiling. To access the detection limits and accuracy of the cDNA microarray raw fluorescence data (10) were plotted versus transcript copy number per cell as determined by kRT-PCR assay for 275 yeast transcripts, Fig. 4a. These genes include: 25 relatively high abundance mRNAs coding for enzymes involved in glycolysis, gluconeogenesis, ethanol synthesis, and glycerol metabolism, previously quantitated by kRT-PCR (5); 65 transcripts encoded by genes on the left arm of chromosome III; and the 185 transcription factor transcripts described above. Both sets of data were for early log phase cultures of a derivative of strain S288C grown in YPD medium (11). The plot recapitulates a second order hybridization curve where most of the change in fluorescence occurs in diploid cells was not detected using the kRT-PCR assay. Expression of the regulatory genes at the MAT locus was very similar for the strains derived from S288C or D27310B. Unexpectedly, expression of a1 and a2 in diploid cells was 4- to 5-fold lower than in mating type a haploid cells or mating type a haploid cells, respectively.

Hybridization-based cDNA microarray and high density oligonucleotide array technologies are widely used for transcript profiling. To access the detection limits and accuracy of the cDNA microarray raw fluorescence data (10) were plotted versus transcript copy number per cell as determined by kRT-PCR assay for 275 yeast transcripts, Fig. 4a. These genes include: 25 relatively high abundance mRNAs coding for enzymes involved in glycolysis, gluconeogenesis, ethanol synthesis, and glycerol metabolism, previously quantitated by kRT-PCR (5); 65 transcripts encoded by genes on the left arm of chromosome III; and the 185 transcription factor transcripts described above. Both sets of data were for early log phase cultures of a derivative of strain S288C grown in YPD medium (10). The plot recapitulates a second order hybridization curve where most of the change in fluorescence occurs between 1 and 100 transcript copies per cell. Below 1 transcript per cell, the fluorescence data are scattered, and the large changes in transcript copy number per cell as measured by kRT-PCR assay are not evident in the raw fluorescence data. A similar curve is obtained when expression levels measured by high density oligonucleotide arrays for log phase cultures of a derivative of strain S288C grown in YPD medium (11) were plotted versus transcript copy number per cell determined by kRT-PCR assay (Fig. 4b). An exponential curve fit to these latter data displays a breakpoint at about 2 transcript copies per cell (Fig. 4c). Transcript copy numbers per cell, determined by kRT-PCR, are in reasonable agreement with both array analyses down to 2 copies per cell. Below 2 transcripts per cell,
However, there is little coincidence between raw fluorescence determined by cDNA microarray or expression level determined by high density oligonucleotide array and transcript copy number per cell determined by kRT-PCR. Furthermore, transcripts displaying significant deviation (scatter) from the exponential curve fit for the cDNA microarray data (Fig. 4a) are different from those displaying significant deviation from the exponential curve fit for the high density oligonucleotide array data (Fig. 4b).

**DISCUSSION**

The results presented here using kinetically monitored RT-PCR extend previous hybridization complexity (1) and SAGE (2) analyses to reveal the full range of transcript abundance in yeast. Yeast transcript abundance ranged from a few hundred copies per cell for glycolytic mRNAs to one-thousandth transcript per cell for transcripts encoding some of the transcription factors. No transcripts were detected from the silent mating type loci in haploid cells or a1 transcript levels were quantitated in strains S273-10B, S273-29A, and S273-6BX29A, which are derived from the parental strain S273-10B. Mating type-specific transcripts were also quantitated in strains YPH499, YPH500, and YPH501, which are derived from the parental strain S288C.

The physiological importance of low level transcripts is illustrated by the a1 transcript (0.003 and 0.008 transcript/cell in the S288C and D273-10B a haploid strains, respectively), which encodes a transcription factor required for expression of a-specific genes (3). The levels of transcripts encoded by IXR1, GAL1, GLN3, SWI6, and SPT3 (0.0009, 0.002, 0.004, 0.004, and 0.005 transcript/cell, respectively) are comparable with or below those observed for the telomere proximal ORFs described above for chromosome III. Remarkably, the level of the IXR1 transcript is comparable with the level of “readthrough” transcript observed for the intergenic region between the divergently transcribed yeast GAL1 and GAL10 genes (0.00075 copy per cell) (5). This latter steady-state concentration would be obtained if a transcript with a half-life of 10 min were synthesized once during a 2-h cell division cycle. Because cellular proteins are typically more stable than the mRNAs that encode them and because the cytoplasm and nucleoplasm are shared during cell division, very low steady-state levels of mRNA can direct synthesis of physiologically important proteins. IXR1 encodes a high mobility group domain protein that binds cisplatin-DNA adducts and represses transcription of the
yeast COX5b gene (14). GAT1 and GLN3 encode members of the GATA transcription factor family involved in activation of transcription of genes subject to nitrogen catabolite repression (15). The SWI6 gene product is involved in regulation of transcription at the G1/S boundary of the mitotic cell cycle (16). Finally, SPT3 encodes a subunit of the SAGA complex involved in histone acetylation (17).

The kRT-PCR assay and the hybridization-based array technologies were compared with respect to transcript detection and quantitation of the 250 yeast transcripts reported here and 25 abundant transcripts previously quantitated by kRT-PCR (5) using published array data for the same transcripts (10, 11). The comparison reveals some significant limitations for array-based detection and quantitation of yeast transcript present at one or fewer copies per cell. The analysis shows that these limitations are imposed by the second order nature of hybridization kinetics. Whereas the array technologies offer the advantage of relatively high throughput analysis of large numbers of transcripts, kRT-PCR offers clear advantages for monitoring mRNA levels over the complete range of intracellular levels. This is an important consideration because the majority of yeast transcripts are present at one or fewer copies per cell. These detection limits are not restricted to yeast, however, because the intracellular range of transcript abundance in bacteria as well as metazoans is likely to be comparable with yeast when adjusted for cell volume and RNA content.

The sensitivity of the kRT-PCR assay is ideally suited for assessing whether or not computationally annotated ORFs are expressed. More generally, kRT-PCR measures the full range of change in transcript level in different genetic, physiological, or developmental changes whereas the array technologies are likely to underestimate the magnitude of such changes. This latter point is illustrated for yeast genes that are subject to carbon catabolite repression. Derepression of the yeast phosphoenolpyruvate carboxykinase (PCK1), fructose bisphosphatase (FBP1), and iso-1-cytochrome c (CYC1) genes measured by kRT-PCR for cells grown on a nonfermentable carbon source versus glucose were 1000-, 200-, and 40-fold, respectively (5). Derepression of these same genes measured by cDNA microarray after diauxic shift (the switch from growth on glucose to nonfermentative growth on the products of glycolysis, ethanol and glycerol) were 14-, 13-, and 3-fold, respectively (10). The levels of these transcripts are above 1 copy per cell after diauxic shift; however, the fully repressed levels of these mRNAs are well below the detection limit of the cDNA microarray assay. Thus, large changes in gene expression can be underestimated or entirely missed by microarray assay depending on the abundance range over which a particular transcript varies.

Acknowledgments—I thank Teresa Yokoi and John Kang for designing the oligonucleotide primers used here and Robert Watson and David Gelfand, Roche Molecular Systems, Inc., for valuable advice and discussion. Recombinant Tth DNA polymerase and uracil N-glycosylase were provided by Roche Molecular Systems, Inc.

REFERENCES
1. Hereford, L. M., and Rosbash, M. (1977) Cell 10, 453–462
2. Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E., Jr., Hieter, P., Vogelstein, B., and Kinzler, K. W. (1997) Cell 88, 243–251
3. Kang, J. J., Kayser, G. A., Jones, H., and Holland, M. J. (1997) Methods 13, 437–443
4. Kang, J. J., and Holland, M. J. (1999) in PCR Applications: Protocols for Functional Genomics (Innis, M. A., Gelfand, D. H., and Sninsky, J. J., eds) pp. 429–444, Academic Press, San Diego, CA
5. Kang, J. J., Watson, B. M., Fisher, M. E., Higuchi, R., Gelfand, D. H., and Holland, M. J. (2000) Nucleic Acids Res. 28, e2 (nar.oupjournals.org)
6. Hinnebusch, A. G. (1997) J. Biol. Chem. 272, 21661–21664
7. Hernandez, N. (1993) Genes Dev. 7, 1291–1308
8. Pinto, I., Ware, D. E., and Hampsey, M. (1992) Cell 68, 977–988
9. Hampsey, M. (1998) Microb. Mol. Biol. Rev. 62, 465–503
10. DelRisi, J. L., Iyer, V. R., and Brown, P. O. (1997) Science 278, 680–686
11. Holstege, F. C., Jennings, E. G., Wyrick, J. J., Lee, T. I., Hengartner, C. J., Green, M. R., Golub, T. R., Lander, E. S., and Young, R. A. (1998) Cell 95, 717–728
12. Laurentson, P., and Rine, J. (1992) Microbiol. Rev. 56, 543–560
13. Herskowitz, I. (1988) Microbiol. Rev. 52, 536–553
14. McAuliff, M. M., Whitehead, J. P., and Lippard, S. J. (1996) Biochemistry 35, 6099–6109
15. Cunningham, T. S., Rai, R., and Cooper, T. G. (2000) J. Bacteriol. 182, 6584–6591
16. Koch, C., Moll, T., Neuberg, M., Ahorn, H., and Nasmyth, K. (1993) Science 261, 1551–1557
17. Grant, P. A., Duggan, L., Côté, J., Roberts, S. M., Brownell, J. E., Caudau, R., Obba, R., Owen-Hughes, T., Allis, C. D., Winston, F., Berger, S. L., and Workman, J. L. (1997) Genes Dev. 11, 1640–1650