# Q-RT-PCR: data analysis software for measurement of gene expression by competitive RT-PCR

Peter A. Doris and Jon K. Hays, Sr.*

Dept. of Cell Biology and Biochemistry and

*Dept. of Computer Services

Texas Tech University Health Sciences Center

Lubbock, Texas 79430

Peter A. Doris, Ph.D.

Dept. of Cell Biology and Biochemistry

Texas Tech University Health Sciences Center

Lubbock, TX 79430

email: cbbpad@wpoffice.net.ttuhsc.edu

**Contents:**

**ABSTRACT.**

We have developed software to assist in the computation of gene expression from data obtained in competitive RT-PCR. Here we describe the mathematical basis of competitive RT-PCR and discuss the criteria which must be met to permit accurate estimations of gene expression to be obtained using this technique. The software which has been developed assists in both the assessment of assay performance and in the routine analysis of data obtained in either titration-based or single tube quantitation of gene expression by competitive RT-PCR. The software is a 100Kb module which functions as a Microsoft Excel add-in. It is compatible with both Windows and Mac versions of Excel 5 and Excel 7 on the Windows 95 platform and employs the spreadsheet, statistical and graphing capabilities incorporated into Excel.

## INTRODUCTION

Quantitation of gene expression by competitive RT-PCR has been developed to the extent that it can now be considered, if applied rigorously, a well-evaluated, robust, quantitative method capable of providing accurate and precise estimates of the abundance of specific transcripts in RNA preparations, even in very small sample sizes (1). However, the mathematical concepts involved are complex and have not been conducive to the rigorous application of this technique. The software we describe here will, it is hoped, remove the problems associated with complex computations and allow the user to evaluate their competitive RT-PCR systems thoroughly and to perform routine data analysis easily.

There are several requirements which can provide for quantitative accuracy in competitive RT-PCR. Many of these have been discussed in our papers (see references below). Some issues (RT efficiency differences between competitors and native template) have been studied and will be reported in forthcoming papers from our laboratory, please check Medline.

The following assumptions are made in using this software:

1. The competitor used is a highly homologous RNA. We have not demonstrated that RT and PCR efficiencies can be identical for heterologous competitors (e.g. Clontech mimics).

2. The user has performed estimates using known starting quantities of both native and competitor RNA and has demonstrated that the system accurately estimates the known native input.

3. The user has performed estimates using known starting quantities of both native and competitor DNA and has demonstrated that the system accurately estimates the known native input.

4. The reaction products are analyzed in a manner which recognizes and reports accurately the amount of heteroduplex products formed in competitive reactions. In our experience, analysis of reactions by denaturing HPLC (see our papers for details) is the most reliable method for such analysis. In addition, except for the initial hardware cost, the method is extremely inexpensive (about 50 cents per analysis) and is also rapid and accurate.

## THEORY

The development of this quantitative method has required that a mathematical approach to competitive reactions be devised and rigorously applied. This approach is founded on mathematical descriptions of the amplification of templates in reactions in which two templates are being simultaneously amplified (4). Amplification of a single template can be described by the equation:

Eqn 1. $$N = N_0 (1+E)^n$$

Where N is the final amount of reaction product, $N_0$ is the initial amount of DNA in the reaction, n is the number of cycles and E is the efficiency of the reaction. If variations in reaction efficiency of only 5% occur between two samples being compared in two different reaction tubes, changes of product amount of more than 100% after 30 amplification cycles will result and make comparison of gene expression levels between the samples meaningless. Competitive RT-PCR is designed to overcome this limitation. The idea is to amplify not only the cDNA of interest in the PCR reaction, but also another cDNA template (competitor) which can be amplified by exactly the same primers. If the amount of the competitor cDNA in the starting sample is known or is always constant between samples it can be used as a quantitative reference providing that it shares necessary similarities with the native template for which it is providing reference. Use of RNA competitors to generate native and competitor cDNA's simultaneously in the same reaction tube permits this technique to be used for accurately quantitating gene expression.

Equation 1 can then be extended to describe the amplification of both templates. If the initial unknown amount of a gene U in a competitive RT-PCR reaction is $U_O$ and that of its specific competitor RNA is $C_O$ and these are subject to *n* reaction cycles in which the efficiency of amplification is $E_u$ and $E_C$ for the unknown and competitor respectively, then from equation 1 (above) we can describe the amount of reaction products at the end of *n* cycles by:

Eqn. 2. $U_n = U_o \cdot (1 + E_u)^n$

Eqn. 3 $C_n = C_o \cdot (1 + E_c)^n$

Making a ratio of Eqn. 2 and Eqn 3. and taking the logarithm gives:

Eqn. 4 $\log(U_n/C_n) = \log U_o - \log C_o + n \cdot \log[(1 + E_u)/(1 + E_c)]$

A basic assumption of competitive RT-PCR which we have recently proven for systems following our own design is that Eu and Ec remain equal throughout the reaction. If true, Eqn. 4 reduces to:

Eqn. 5 $\log(U_n/C_n) = \log U_o - \log C_o$

In calculating competitive RT-PCR reactions to obtain the unknown amount of a gene (Uo) present in a sample, a plot is made which relates log(Un/Cn) to logCo, the known amount

of starting competitor RNA. This allows logUo to be calculated. The estimated initial value for amount of native gene expressed in a sample (Uo) is the antilog of this value.

Eqn. 5 indicates that such a plot will form a straight line having a slope of -1 (or of 1 if log(Cn/Un) is plotted) because it is of the form $y = ax \pm b$ where the value of a must be 1 or -1 to follow the model.

The calculations to be performed in establishing a competitive RT-PCR system and demonstrating that it is in concordance with these mathematical constraints are complex. However, once such a system has been established and accordance with mathematical requirements has been demonstrated, the amount of unknown gene expressed in a sample can be estimated in single tube assays, removing the need for titrations, once the approximate abundance has been determined. However, even with such a simplified system, estimates of initial abundance of the target template require difficult mathematics in which error can be easily introduced. For this reason, we have developed software to aid both in establishing the initial performance characteristics of a competitive RT-PCR system and also to provide routine data analysis so as to estimate the abundance of a template based on measurements of reaction product accumulation and several other parameters describing the reaction input and products.

Our own work in developing competitive RT-PCR systems has shown that the use of homologous competitors is essential if the competitive system is to produce accurate estimates (2, 3). Homologous competitors, however, have the capability of producing heteroduplex products formed by the annealing of the competitor amplicons with highly homologous native amplicons. In our experience, such heteroduplexes are usually not

resolved by non-denaturing agarose gel electrophoresis. This may lead to the erroneous assumption that, since heteroduplexes are not visible by this technique, they are not present. The introduction of a novel, inexpensive and highly accurate PCR product analysis technique based on ion pair reversed phase HPLC has resolved the important question of heteroduplex formation and permits accurate resolution of both homoduplex and heteroduplex reaction products (2, 3). In light of the insight gained in these experiments, it seems safest to assume that heteroduplexes are always present at the conclusion of competitive reactions unless positive evidence to the contrary exists. Since the mathematical method involves analysis of reaction product ratios and heteroduplex formation disturbs the final homoduplex product ratios, it is essential that heteroduplexes be identified, accurately quantitated and that their contribution to final product ratios be included in the analysis.

## SOFTWARE REQUIREMENTS

The data analysis system we have devised was produced using the visual basic programming tools in Microsoft Excel. The resulting program, Q-RT-PCR, has been compiled as a Microsoft Excel add-in module (q-rt-pcr.xla). When Excel is installed, it is important to perform a complete installation. This ensures that other add-in modules are available to the program.

## INSTALLATION

*Install Excel Data Analysis Tools*

Go to the Tools Menu in Excel, pull this menu down and check to see whether "Data analysis" is listed at the bottom of the menu. If not, select "Add Ins" from the Tools menu and click on the box in the following dialog next to "Analysis ToolPak".

Data Analysis should now be available under the tools menu. If it is not, Q-RT-PCR will not run, but will present a dialog indicating that this installation needs to be completed.

*Install Q-RT-PCR*

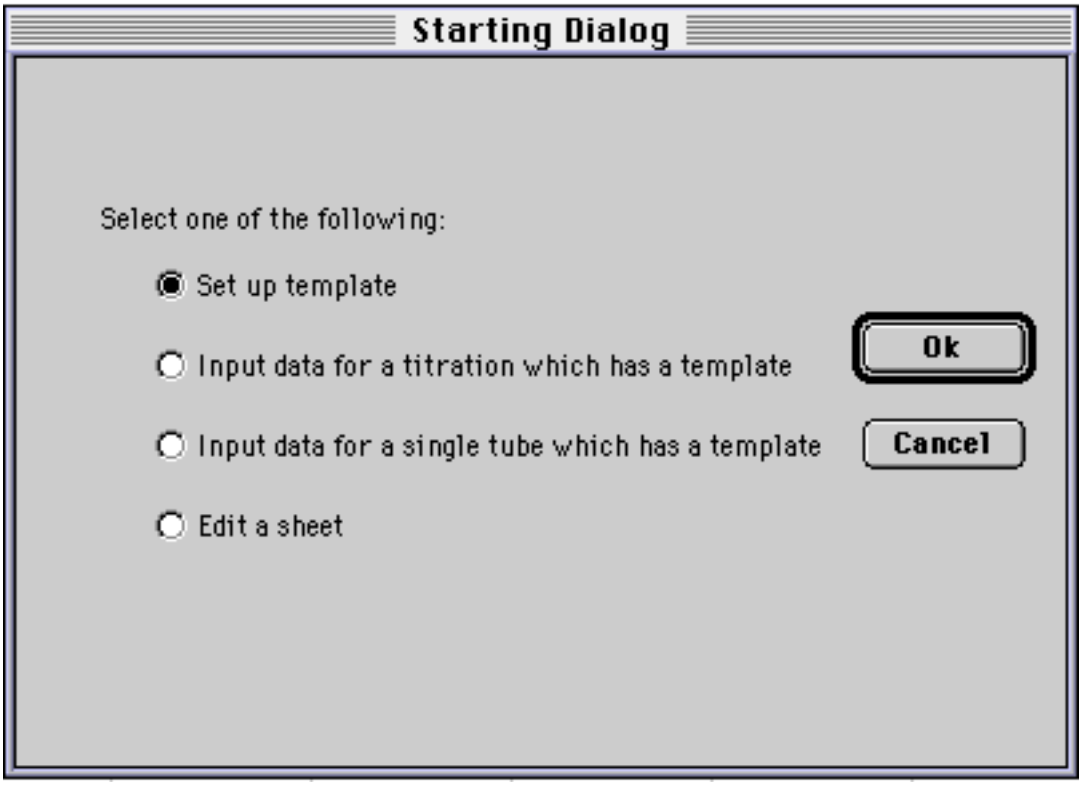
After downloading q-rt-pcr.xla, this file should be moved into the Library directory inside the Microsoft Excel installation directory (this is called Macro Library on the Mac platform), you may find this inside the Microsoft Office directory. Go to the Tools menu and select Add Ins. Check the box in the following dialog next to Q-RT-PCR. On completion, Q-RT-PCR and Data Analysis should both be available under the Tools menu. In future, you may notice in the lower left of the Excel window that q-rt-pcr.xla is loaded

as Excel starts up. An information dialog describing Q-RT-PCR will be displayed, click on OK or press Enter (press Return on Mac) after reading this dialog. You are now ready to begin using the software.

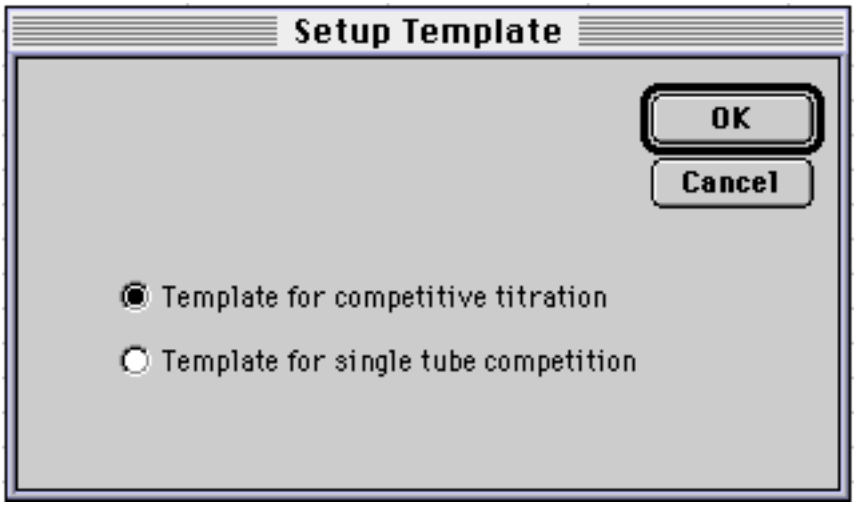
## USE

*Running Q-RT-PCR.*

Select Q-RT-PCR from the Tools menu of Excel. The program presents a window like the one below:

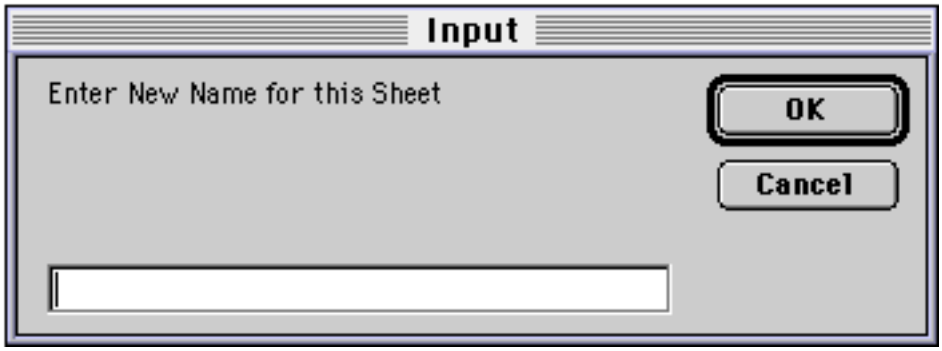



During the first use, it is necessary to set up a template to provide the software with basic information concerning your assay system. This template can be for titration analysis or single tube analysis

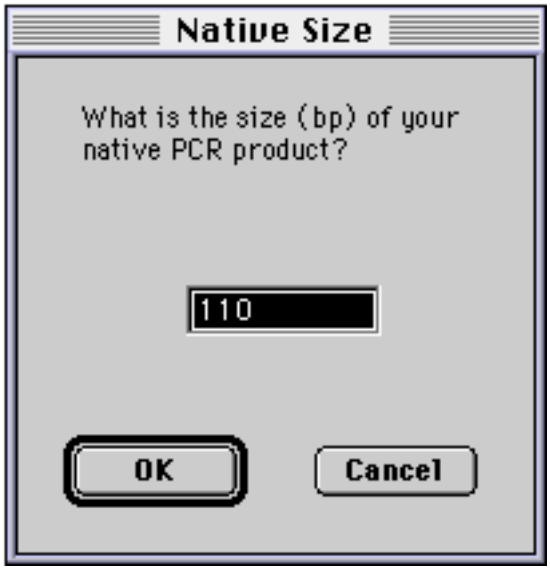



Titration analysis must always be performed first in order to assure that the assay system is meeting mathematical ideals discussed under **THEORY**, above. Titration of increasing amounts of competitor against a constant amount of standard will result in a competition in which reaction product ratios should shift from predominantly the more abundant native template (at low doses of competitor) to predominantly the more abundant competitor template (at higher doses of competitor). If both templates are amplified with identical efficiency throughout the competitive PCR reaction, the resulting titration will produce a straight line with a slope of unity. (Note, this does not mean that RT efficiency is identical for both templates, this must be examined separately to demonstrate a truly accurate system.)

The next dialog requests that you give a name to the worksheet that Excel will now create, this name will appear on the tab at the lower edge of the worksheet.



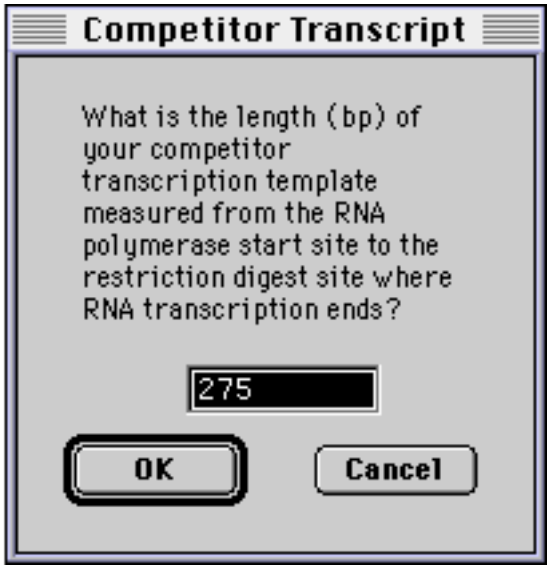

For example, you might call the worksheet: GAPDH titration template. This will be followed by a dialog which asks you to enter the size of the native PCR product in base pairs, you may modify the number in the window from the keyboard, only numeric characters may be entered:




Then you will be asked to enter the size of the competitor product in base pairs:


Again, the numeric value can be changed to suit your system.

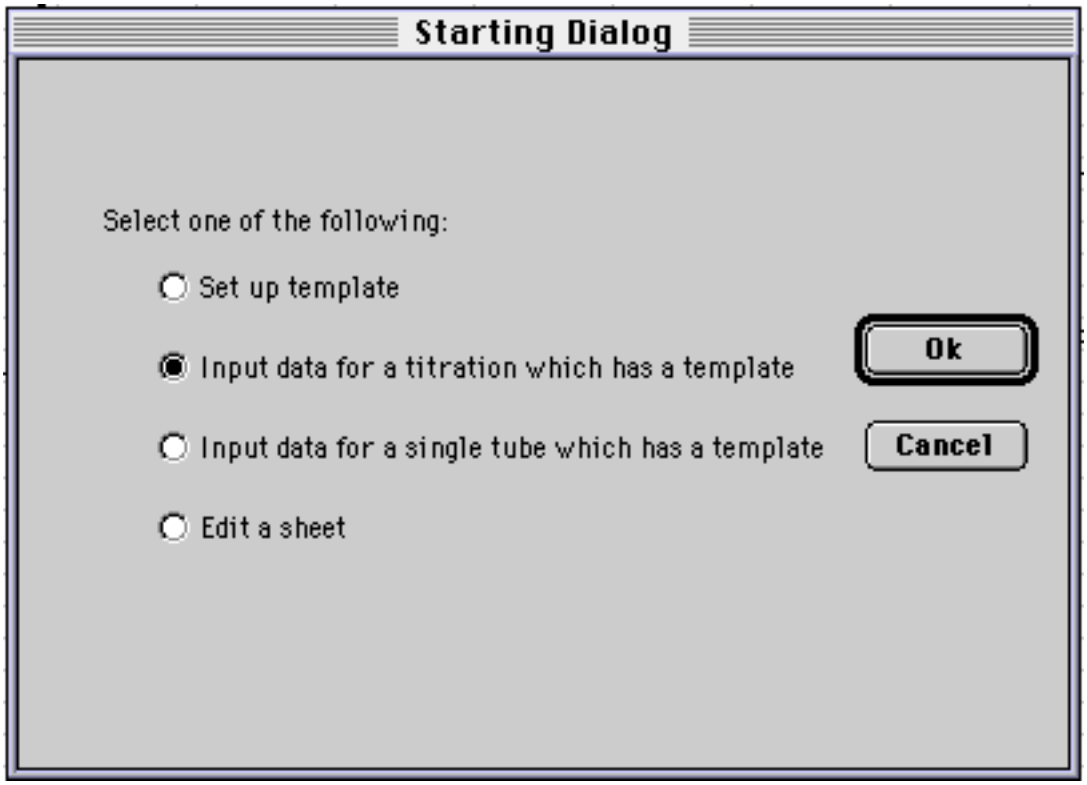
Finally, you will be asked to indicate the length of the competitor RNA in base pairs. This will almost certainly be different from the size of the competitor PCR product. Your competitor RNA will presumably be an in vitro transcript, probably synthesized from a plasmid containing the competitor cDNA and an RNA polymerase promoter. You will restriction digest the plasmid downstream in the cDNA and perform run-off transcription. Presumably you will purify the run-off transcribed RNA and measure its concentration by UV absorbance. You will then estimate the amount of competitor RNA added to each reaction from this information (e.g. 10 fg). Information on the size of the competitor transcript is used by the program to estimate the number of molecules of native transcript present in the sample.




At this point, a named worksheet is completed. In order to use it, you will return to the Tools menu and select Q-RT-PCR once again. This sequence of events would have been no different if you had selected set up of a template for single tube assay. The next step will run through data entry and analysis in titration and single tube templates

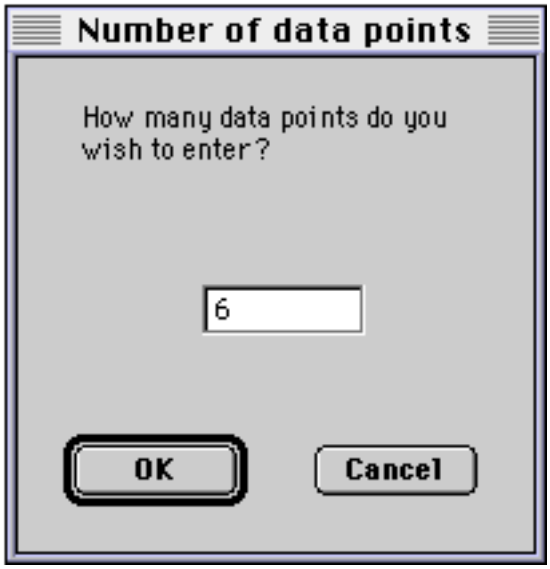
*Data entry and analysis (titration).*

After selecting Q-RT-PCR from the Tools menu again, you will once again be presented with the Starting Dialog:



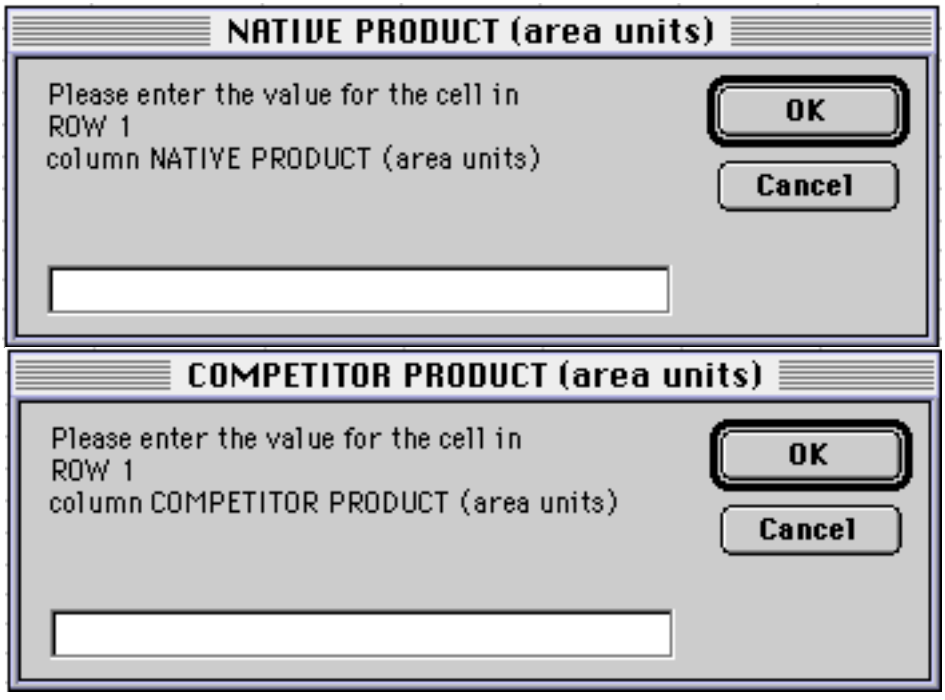

Select (as shown above) the second radio button. The next dialog requests that you identify the worksheet which will server as the template for data entry. You will select this worksheet from the drop down list which appears in the dialog window. It will be a worksheet you have created during a previous set up routine.

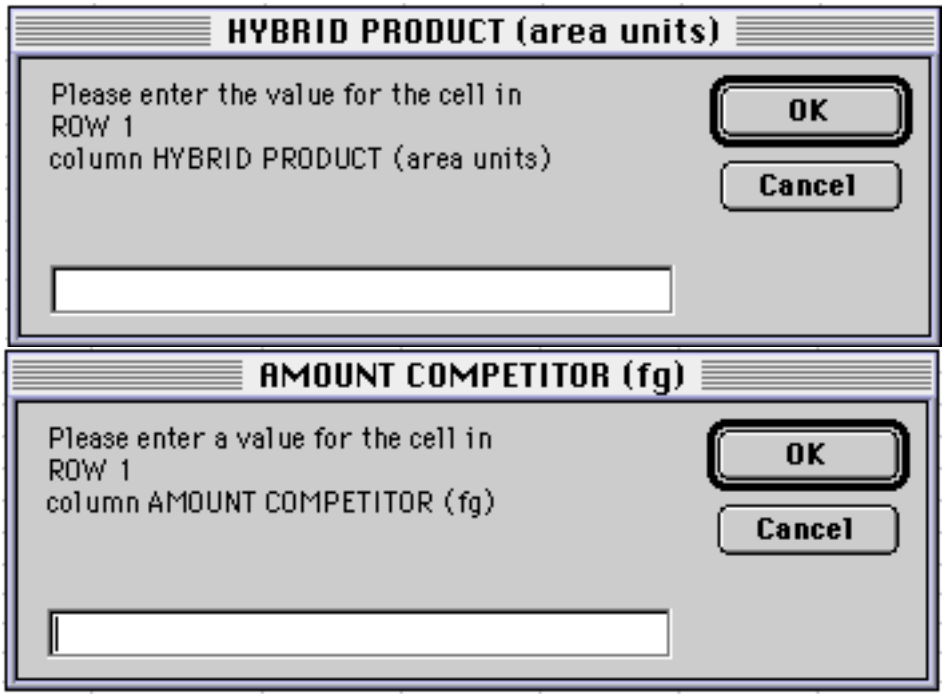

You will now be asked how many data points form your titration:




Enter the appropriate number in the range 3-9.

The next series of dialogs permits sequential entry of data for each reaction. Four dialogs are required to enter the data for each reaction. You will loop through them until all data has been entered. These dialogs are, in sequence:





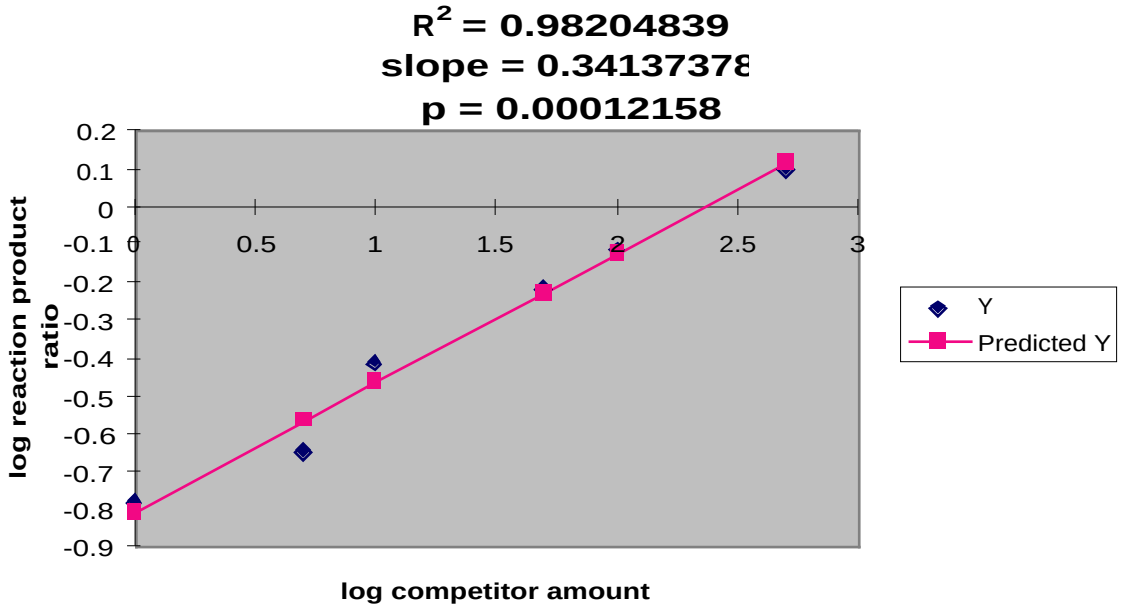

You will be prompted after each to verify that the numbers you have entered are as you intended them to be, e.g.:


These dialogs can be tedious, but it is easier to make corrections now, than after the spreadsheet has been built.

At the conclusion of data entry, the spreadsheet contains the entered values as well as the estimated unknown (see example below).

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 55024 | 12362 | 14332 | 59403 | 22314.8 | 1 | 9818.5 | 0.16529 | -0.78176 | 0 | | | | |
| 48993 | 15662 | 17645 | 54385 | 27915.5 | 5 | 12282.8 | 0.22585 | -0.64618 | 0.69897 | | | | |
| 36590 | 22354 | 23458 | 43758 | 38644.3 | 10 | 17003.5 | 0.38858 | -0.41052 | 1 | | | | |
| 25231 | 26984 | 29872 | 34359 | 47728.4 | 50 | 21000.5 | 0.61122 | -0.2138 | 1.69897 | | | | |
| 21776 | 34225 | 26734 | 29945 | 52790.3 | 100 | 23227.7 | 0.77569 | -0.11031 | 2 | | | | |
| 14765 | 46982 | 21443 | 21317 | 61873 | 500 | 27224.1 | 1.27711 | 0.10623 | 2.69897 | | | | |
| | | | 0 | 0 | | | | | | | | | |
| | | | 0 | 0 | | | | | | | | | |
| | | | 0 | 0 | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | 0 | | 2.33542 | 216.482 | 1397030 | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| NATIVE PR | COMPETI | HYBRID PI | correc nati | correct con | AMOUNT | finalcorrect | ratio | | logratio | logstandard | logunknown | amountunk | amount unknown |
| area units) | (area units) | (area units) | (area units | (area units) | (fg) | (area units) | | | | | | (fg) | molecules |

An additional worksheet is created by the automatic running of the Excel Data Analysis add-in which performs regression analysis of the titration and displays the regression statistics and regression plot. The sheet containing this information has the same name as the sheet into which you have just entered data, except the suffix .stat is added.

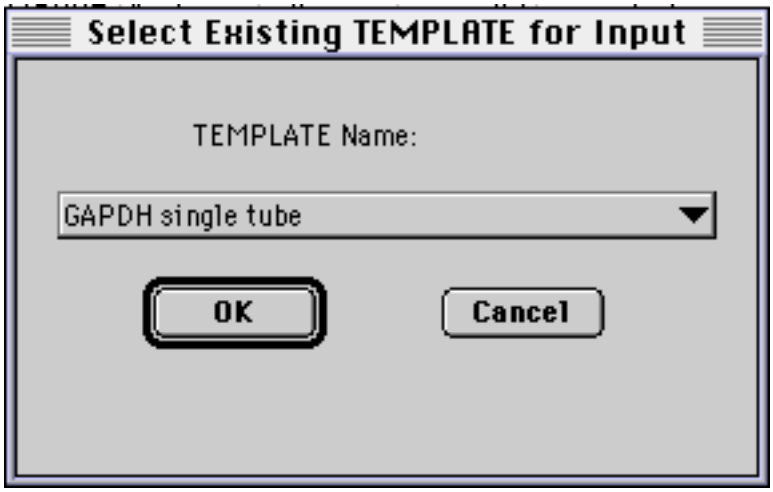



*Data entry and analysis (single tube assays)*

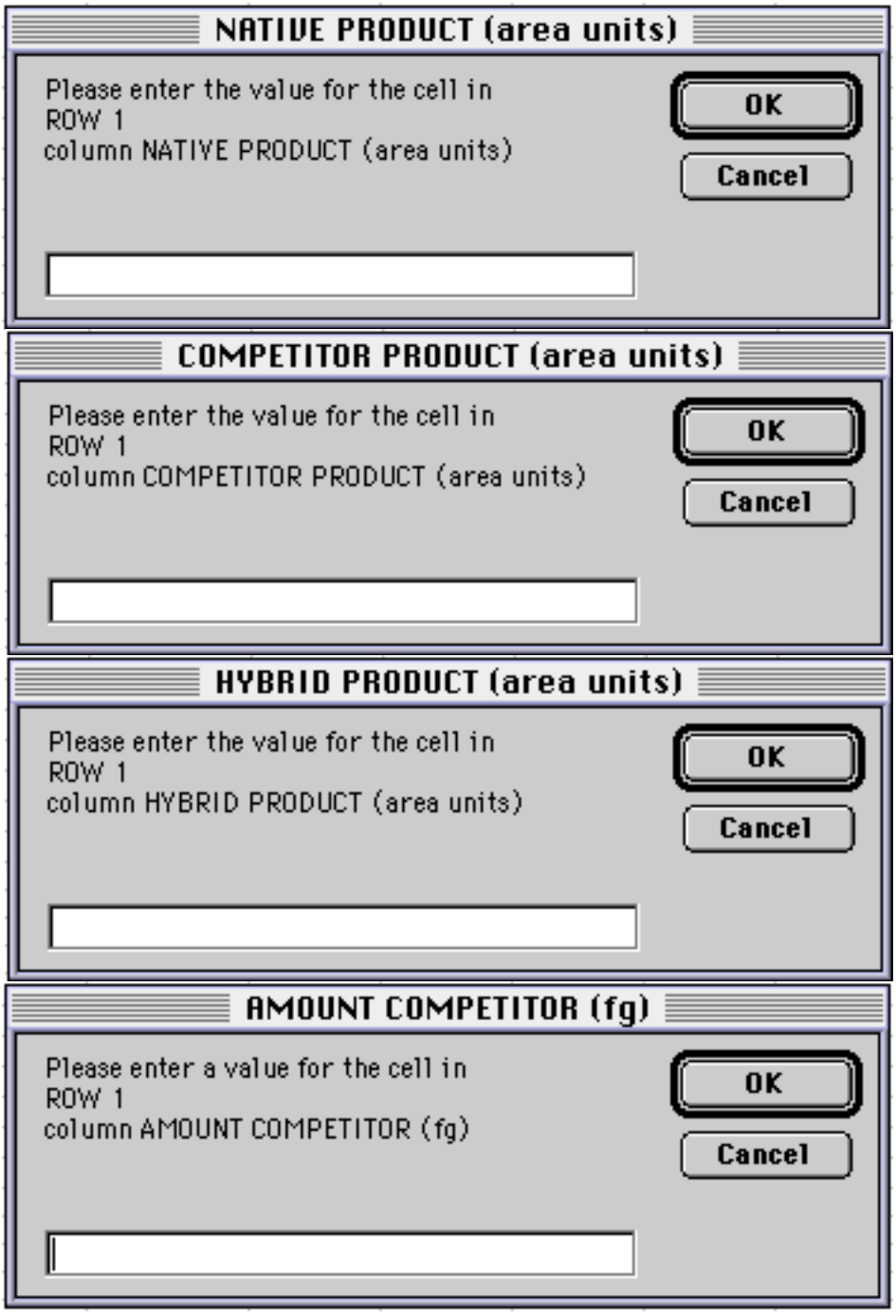
Competitive RT-PCR systems which have been shown on titration to generate lines which are straight and have a slope of unity can be considered reliable for quantitation (4). Once shown, quantitation can be simplified so that only a single reaction is needed to quantitate

each sample (assuming that the competitor dose and native level of gene expression are roughly equimolecular. This is true because the ratio of reaction products at the end of the reaction will always be the same as the ratio of reaction products at the beginning of the PCR reaction if the amplification efficiencies are identical and if they decline identically during the reaction. Thus, since the final reaction product ratio can be estimated from the reaction products and the amount of one of the initial reaction templates (competitor) is known, the unknown can be determined without titration.

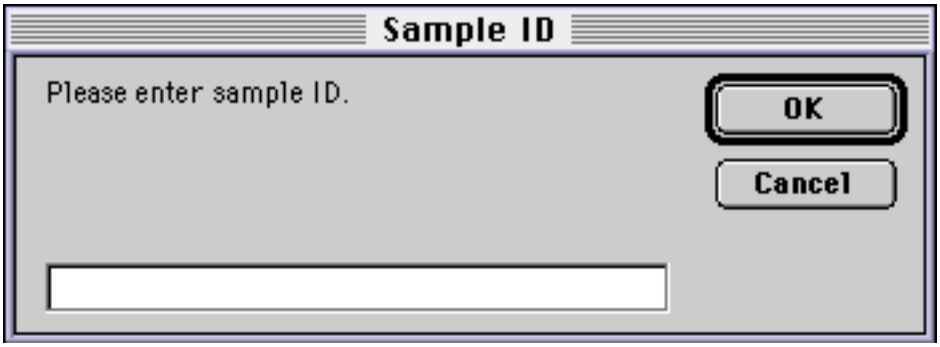
After selecting Q-RT-PCR from the Tools menu again, you will once again be presented with the Starting Dialog, Selecting the button labeled “Input data for single tube which has a template” will result in a new dialog window containing a drop down list of available sheets (including any single tube template you have already created and saved), for example:




You will now be presented with a series of four dialogs to enter the data for each sample. This data will complete a single row of the spreadsheet. These are the same dialogs as used for data entry in titration analysis:




At the end of this cycle of data entry you will requested to enter information concerning the sample ID:



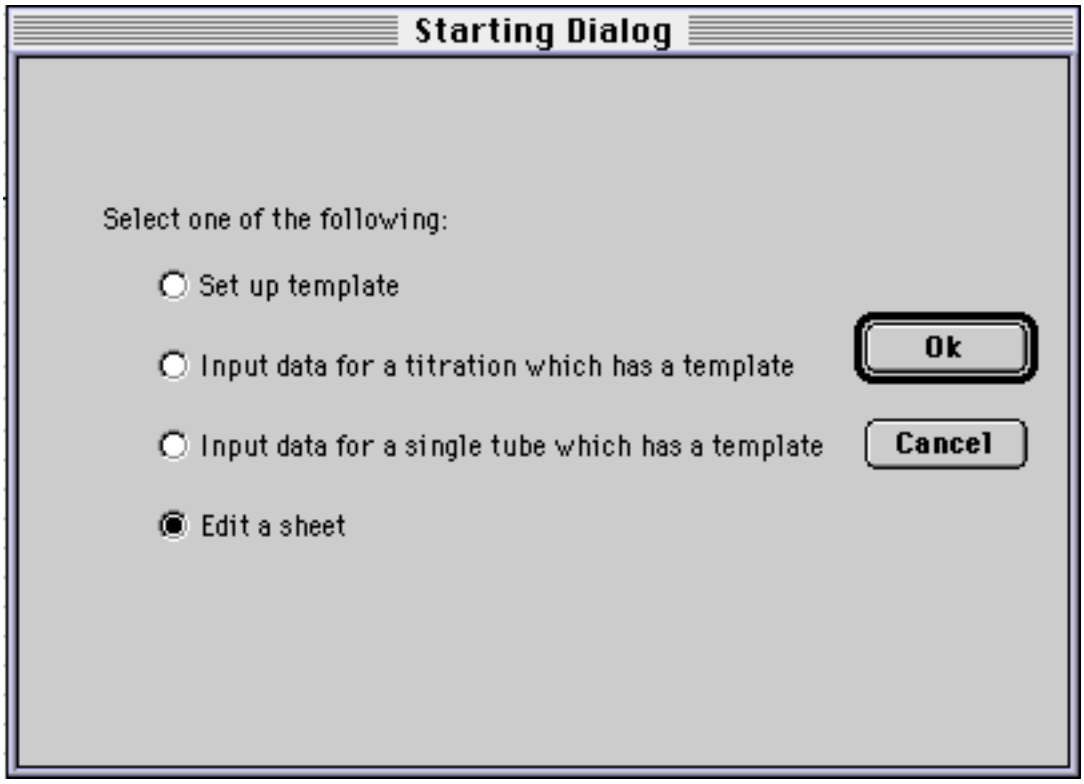

In order to enter data from multiple samples a dialog appears asking whether you have more data to enter:


If you click Yes, the data entry loop will repeat, this time entering data on the second row of the spreadsheet. If you click No, data entry will end and the spreadsheet is complete. A completed spreadsheet is shown below. The calculated unknown value (in molecules s indicated in the column to the left of the sample ID. The unknown amount is also indicated in fg in the next left row.

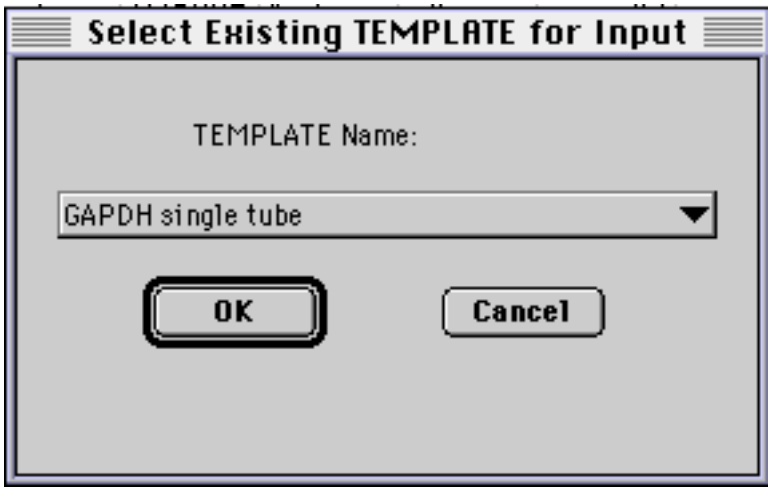
| 1000 | 100 | 200 | 1061 | 238.88889 | 10 | 105.111 | 0.09906 | 0.99058 | 6392.53 | | Heart mouse1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1782 | 345 | 533 | 1945 | 715.13889 | 10 | 314.661 | 0.16179 | 1.61791 | 10440.9 | | Heart mouse 2 |
| 8733 | 5423 | 5548 | 10428 | 9275.7778 | 10 | 4081.34 | 0.39137 | 3.91375 | 25256.7 | | Heart mouse 3 |
| 8801 | 6522 | 6621 | 10824 | 11119.917 | 10 | 4892.76 | 0.45203 | 4.52026 | 29170.8 | | Heart mouse 4 |
| 9932 | 6533 | 3487 | 10997 | 8954.5278 | 10 | 3939.99 | 0.35826 | 3.58263 | 23120 | | Heart mouse 5 |
| 8223 | 5422 | 5266 | 9832 | 9078.9444 | 10 | 3994.74 | 0.4063 | 4.06297 | 26219.7 | | Heart mouse 6 |
| | | | 0 | 0 | | 0 | #DIV/0! | #DIV/0! | #DIV/0! | | |
| | | | 0 | 0 | | 0 | #DIV/0! | #DIV/0! | #DIV/0! | | |
| | | | 0 | 0 | | 0 | #DIV/0! | #DIV/0! | #DIV/0! | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| NATIVE PR | COMPETIT | HYBRID PR | correc native | correct compe | AMOUNT ( | finalcorrecti | ratio | answer(fg) | amountunknown | | sample ID |
| area units) | (area units) | (area units) | (area units) | (area units) | (fg) | (area units) | | | molecules | | |

*Data editing*

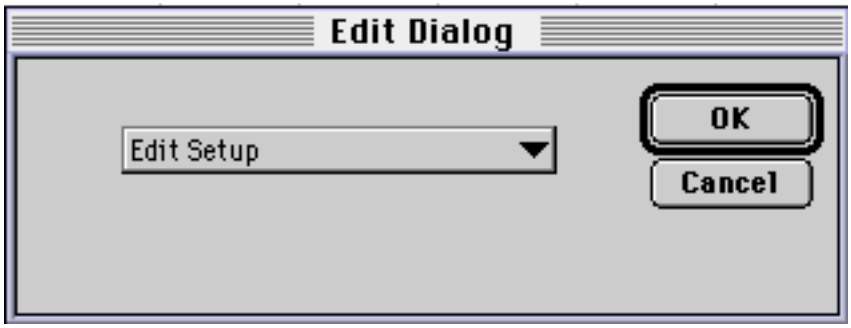
Existing sheets can be edited from the Starting dialog. Go to the Tools menu and select Q-RT-PCR. Click on the edit button as indicated below:

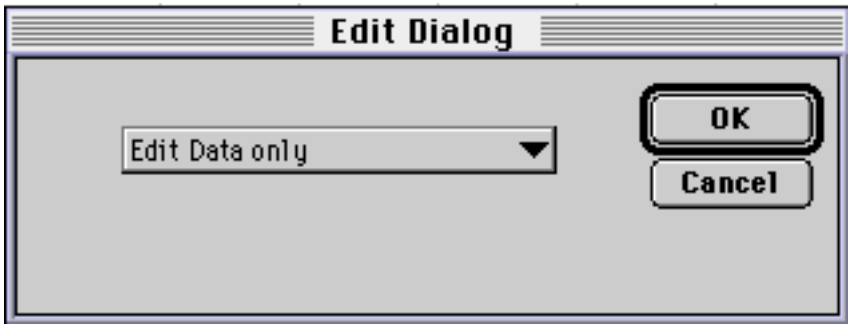



The dialog which follows presents a drop down list of sheets (templates) which you may wish to edit. Select one:



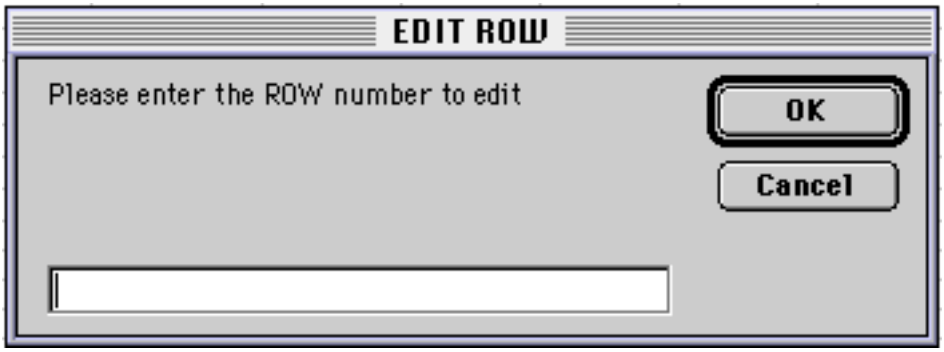
The next dialog allows you to indicate whether you wish to edit the set up parameters (i.e. reaction product sizes, transcript size etc.) or reaction product data you have previously entered, there are two selections in the drop down list which make this possible:




If you choose to edit set up, you will be lead through the dialogs which allow you to enter information on the reaction product sizes and competitor transcript size (base pairs) once again.

If you choose to edit data, you will be asked to re-enter data derived from analysis of the amounts of reaction products accumulated, the dialog will allow you to select which data point (spreadsheet row) you wish to modify:




You may continue in this fashion until all desired modifications are made.

When you edit a titration spreadsheet, the regression analysis is refreshed after editing of each row and a new sheet is created displaying the regression analysis results. You are requested to name this sheet.

## NOTICE

This software is provided without any warranty. It is for research use only. It is expected that individuals using it will be knowledgeable molecular biologists capable of and responsible for verifying through other means that the analysis performed by this tool is accurate. The authors and Texas Tech University Health Sciences Center make no assurances about reliability, accuracy or suitability for the purpose described. Use at your own risk. This software is not intended for clinical or diagnostic use. Software is copyright by TTUHSC. You may not make copies, nor may you distribute copies. Use of this software by FOR-PROFIT ORGANIZATIONS is not permitted unless a license is first secured from TTUHSC. Please contact the Intellectual Property Officer, Office of General Counsel, Texas Tech University Health Sciences Center, Lubbock, Texas 79430 (806) 743 2986 for licensing information and assistance.

## PROBLEM REPORTS

Development of this software is considered a completed project. It is unlikely that further revisions or upgrades will be forthcoming in the foreseeable future. However, if you encounter problems, please email a report to csjkhi@ttuhsc.edu. If possible, we will try to assist.

**REFERENCES.**

1. Hayward-Lester, A., B. S. Chilton, P. A. Underhill, P. J. Oefner and P. A. Doris. 1997. Quantification of specific nucleic acids, regulated RNA processing and genomic polymorphisms using reversed-phase HPLC. F. Ferré. Gene Quantification. Birkhauser, Boston.

2. Hayward-Lester, A., P. J. Oefner and P. A. Doris. 1996. Rapid quantification of gene expression by competitive RT-PCR and ion-pair reversed-phase HPLC. BioTechniques 20:250-257.

3. Hayward-Lester, A., P. J. Oefner, S. Sabatini and P. A. Doris. 1995. Accurate and absolute quantitative measurement of gene expression by single tube RT-PCR and HPLC. Genome Res. 5:494-499.

4. Raeymaekers, L. 1993. Quantitative PCR: theoretical considerations with practical implications. Anal. Biochem. 214:582-5.