MAKERGAUL: An innovative MAK2-based model and software for real-time PCR quantification

Christoph André Bultmann *, Ralf Weiskirchen *

Institute of Clinical Chemistry and Pathobiology, RWTH University Hospital Aachen, D-52074 Aachen, Germany

A R T I C L E   I N F O
Article history:
Received 1 August 2013
Received in revised form 11 October 2013
Accepted 19 October 2013
Available online 30 October 2013

Keywords:
qPCR
Expression analysis
DNA quantification
Computer-based analysis
MAKERGAUL
MAK2
MIQE
Open source software

A B S T R A C T

Objectives: Gene expression analysis by quantitative PCR is a standard laboratory technique for RNA quantification with high accuracy. In particular real-time PCR techniques using SYBR Green and melting curve analysis allowing verification of specific product amplification have become a well accepted laboratory technique for rapid and high throughput gene expression quantification. However, the software that is applied for quantification is somewhat circuitous and needs actually above average manual operation.

Design and methods: We here developed a novel, simple to handle open source software package (i.e., MAKERGAUL) for quantification of gene expression data obtained by real time PCR technology.

Results: The developed software was evaluated with an already well characterized real time PCR data set and the performance parameters (i.e., absolute bias, linearity, reproducibility, and resolution) of the algorithm that are the basis of our calculation procedure compared and ranked with those of other implemented and well-established algorithms. It shows good quantification performance with reduced requirements in computing power.

Conclusions: We conclude that MAKERGAUL is a convenient and easy to handle software allowing accurate and fast expression data analysis.

© 2013 The Canadian Society of Clinical Chemists. Published by Elsevier Inc. All rights reserved.

Introduction

Reverse transcription quantitative polymerase chain reaction (RT-qPCR) is a popular rapid quantification technology with capacity to detect and measure minute amounts of nucleic acids in relative or absolute term in nearly any kind of biological sample. Based on the practical simplicity of this technique, a wide variety of applications of this technology have been developed in research and diagnostic. However, there are several potential experimental drawbacks and shortcomings that might arise from inadequate storage of samples taken for RNA isolation and cDNA synthesis, poor choice of primers used in PCR, occurrence of PCR inhibitors in samples, unidentified contamination, and inappropriate data and statistical analysis. All these factors may potentially result in inadequate and conflicting data [1]. In addition, most often published studies provide incomplete information about the experimental setup.

In addition, appropriate selection of reference genes is known to be important to obtain accurate and reproducible RT-qPCR results [2,3]. It is further known that the examination of raw data per se, evaluation of quality and reliability of measurements, and the generation of reportable (interchangeable) results strongly affect resolution, precision and robustness of a PCR method [4]. In particular, the estimation of the PCR efficiency, variable factors affecting the efficiency value, and the accurate mathematical evaluation by various qPCR calculation models were identified as critical parameters in absolute and high precision DNA quantification [4].

In most laboratories, the standard curve technique for absolute quantification or the $2^{-ΔΔCt}$ method are taken to estimate target gene concentration [5,6]. However, both strategies assume amplification efficiencies to be identical or even at optimum for both the target and reference templates. Therefore, several other models for accurate quantification of qPCR data were subsequently developed. In most models, the shape of a single qPCR amplification curve was proposed to be sufficient to uniquely determine initial DNA concentration [7–11]. Based on the fact that the amplification rate correlates to the amplicon’s quantity, it is in principle possible to allow target quantification via linear regression analysis. Therefore, implementation of simple software packages for absolute quantification of qPCR data is possible [11]. More recently, a two-parameter mass action kinetic model of PCR, i.e., the MAK2 algorithm, was presented that allows accurate quantification of target concentration from a single qPCR assay without construction of standard curves [12].
We here adapted and expanded this elaborated model and integrated it into a novel easy to handle open source software package, i.e., MAKERGAUL, for estimating gene expression data by real time PCR technology. We evaluated the performance of our software with an already well characterized real time PCR data set that was recently published together with all statistical evaluation sheets. Based on our evaluation, we anticipate that the MAKERGAUL software allows rapid data analysis and quantification without the need to generate a standard curve, or the need for normalization to a selected reference gene.

**Material and methods**

**Quantification model**

For quantification of qPCR data, we considered to create a model that allows accurate quantification of target concentration from a single qPCR assay with high computation speed, without the demand to construct a standard curve, or the need for normalization. These requests were realized by combining a mechanistic, two-parameter mass action kinetic model of PCR, i.e., the MAK2 model, with findings of a study demonstrating that the main factor that is responsible for the plateau phase in PCR is caused by binding of DNA polymerase to its ampli-

In brief, the MAKERGAUL algorithm described in this study. The developed model were entered in mathematical formulas and are the basis of the tion products [12,13]. In brief, the model that we have generated is demonstrated to work on samples that contain competimer.

DNA_{cycle n} = DNA_{cycle n-1} + DNA_{cycle n-1} \times \frac{\text{Enzyme usable n}}{\text{Enzyme usable n} + DNA_{cycle n-1}} (1)

Fluorescence_{cycle n} = DNA_{cycle n} + Fluorescence_{baseline} (2)

In formula 1, we defined Factor_{inhibition} as a variable for double stranded DNA (dsDNA) inhibiting the DNA polymerase and inserted it into the MAKERGAUL algorithm (for more details about the analysis model see Supplemental material 1).

**Software development**

For development of a platform-independent, open source and easy to use software that includes our analysis algorithm, we developed a server side solution that we implemented in different programs and server side scripting languages such as PHP [14–16], HTML [17], CSS [18], JavaScript [19], C++ [20–24] including the libraries GMP v. 5.1.2 [25] and MPFR v. 3.1.2 [26], as well as some other helpful components [27,28].

In brief, the user interface is created in and distributed by PHP programs and HTML in the client’s web browser, while data analysis is carried out in C++–based subprograms. This strategy allows in future integrating and linking new analysis tools to the program via new PHP scripts and objects. In addition, in cases of problems with server execution rights the analysis can also be realized in PHP (for more information refer to Supplemental material 2). The user interface is built in a classic “plate setup screen” that allows reading of fluorescence data, individual wells, and data processing with the analysis module. The final software is enclosed in Supplemental material 3 that further contains the file “readme.txt” providing some important safety precautions for installing and use of our software.

**Model evaluation**

For evaluating the performance of our model, we re-analyzed an existing, well documented RT-PCR data set developed by Vermeulen et al. [29]. Concisely, the complete collection of gene expression data contains fluorescence values from 64 different plates with samples from patients and gene specific standard dilution series that were generated on a standard LightCycler480 system (Roche Diagnostics, Mannheim, Germany) with primers for different target genes. This data set was recently used in another large comparative study in which different methods for analyzing amplification curves of RT-qPCR were rigorously compared and referred in a respective study as ‘biomarker data set [30], a term that we will also use in our study in the following. For our evaluation, we only used the four-point 10-fold dilution series of each plate. In addition, the data set of “AluSq” was excluded because the developed algorithm is not capable to work on samples that contain competimer. In a first step, the computation of each sample was done with the MAKERGAUL model and collected data was inserted into a prepared copy of “analysis_dilution_series.xls” described elsewhere [30]. Using this data sheet we performed an extensive statistics for MAKERGAUL to get information about the performance parameters (absolute) bias, precision, resolution, linearity and changed variability (details see Supplemental material 4). Based on our analysis, we found that the parameter Factor_{inhibition} varied only in a small range. To understand its influence on the outcome of our analysis we added an additional software module termed MAKERGAUL_C that differs to the original MAKERGAUL component in its possibility to assign Factor_{inhibition} a fixed value. Using this module we then re-analyze the data set, using the mean of the calculated Factor_{inhibition} from every dilution series as a fixed value in the subsequent cycle.

To allow comparison of the MAKERGAUL and MAK2 models, a version of the MAK2 algorithm described previously [12] was integrated.
as an independent module in our software. The performance of the two models was not separately compared because the C++ version requires the MPPR library additional to the standard arbitrary precision GMP library. As a consequence, this fact would otherwise artificially handicap and worsen the performance and processing time in the MAK2 model.

We next compared in more detail the performance of our two algorithms with the results of the methods examined and described in [30] as they apply to 5PSM [31], Cy0 [32], CAmpER: DART [33], CAmpER: FPLM [34], FPK-PCR [35], LinRegPCR [36], LRE-qPCR [37], MAK2 [12], and PCR-Miner [38]. Since a more detailed analysis of all these algorithms is beyond the scope of this article, we focused on some special characteristics of particular methods. A good overview of all the different quantification methods can be found in the Supplemental material of [30].

To perform individual algorithms with MAKERGAUL and MAKERGAUL_C, we incorporated the original quantification data sets achieved by the methods (given in the supplements of [30]) into prepared copies of “analysis_dilution_series.xls” as described above. Like for MAKERGAUL and MAKERGAUL_C, we only analyzed the data of the four-point 10-fold dilution series of the 63 genes from the original biomarker data set. To obtain consistent implementation-based errors, we further decided to fall back on the original MAK2-delivered DNA values from this data set instead of recalculating them in our software. The respective DNA values that served as gold standard, named ‘Standard-Cq’ [30], were obtained and used with the kind permission of Jan M. Ruijter. For LRE qPCR, there were two data sets available, using a fixed (LRE-E100) and a variable (LRE-Emax) PCR-efficiency for DNA calculation.

The parameters absolute bias, linearity, precision and resolution for the 63 genes were compared by EXCEL and an additional program [39], performing a Friedman test [40] and including the methods mentioned above. The null hypothesis in this statistic test is an equality of all algorithms, expressed as an unordered set of the methods’ ranking per gene for the respective performance parameter analyzed. When the hypothesis was rejected, a multiple comparison of the groups was done to determine which subsets of methods are different [41]. Finally the mean rank considering the individual ranking of the four performance parameters per method were analyzed in a Friedman-test as described above. The complete results of this comparison are shown in Table 1 and Supplemental material 5.

Results and discussion

Reliable detection and quantification of mRNA are fundamental in all areas of molecular biology. There are numerous protocols available that allow amplification of a specific mRNA and quantification of respective target nucleic acid. However, the strategy and algorithms used for data analysis most often requires establishment of a time- and labor-intensive standard curve for each gene investigated. In addition, the software that is applied for quantification is often circuitous and needs above average manual operation. Therefore, we tried to develop a novel, simple to handle, open source software package for estimating gene expression by real-time PCR technology without the need for setting up a calibration curve or requirement to perform a normalization with a housekeeping gene.

Model developing outcome

To establish an appropriate model for our software, we combined the MAK2 model that describes the accumulation of amplicon DNA during PCR [12] with the fact that the main factor contributing to the plateau phase in which the amplification reaction is lowered down consists of DNA polymerase that binds to its amplification products [13]. In the final software we have developed, i.e., MAKERGAUL, we developed the following set of theoretical assumptions: (i) the educts of the DNA polymerase is a DNA single strand and the product a double strand. In each PCR cycle every DNA strand can be maximal doubled; (ii) the generation of novel strands is dependent on the content of free DNA polymerase [13]; (iii) the higher the DNA concentration, the more the DNA polymerase is occupied by the DNA during the cycle, and the less often a DNA strand encounters a free polymerase that copies it; and (iv) DNA polymerases bind with a certain probability also to double stranded DNA [13]. As a consequence, they are no longer available in the replication phase and the amount of usable polymerases decreases in each cycle simultaneously; (v) the measured fluorescence of the real-time PCR corresponds to the DNA concentration, along with the basis of fluorescence in the measurement system [12]; and finally (vi) primer deficiency plays no role in the late phase of the PCR when amplification is lowered down [13]. Based on these assumptions, we have developed mathematical algorithms that were integrated into MAKERGAUL.

One critical factor in all these calculations is the circumstance that the precise factor at which each cycle is inhibited by binding of DNA polymerase to double stranded DNA (dsDNA) is not known. Nevertheless, we thought that the definition of Factor inhibition as a constant instead of a variable would enhance the overall accuracy of DNA quantification. For this reason, we next determined the mean values for inhibition of each analyzed gene with MAKERGAUL using again the published biomarker data set of Ruijter et al. [30]. Although all these values were only marginally different in each set, they slightly varied between the different genes. This finding demonstrated that Factor inhibition is an amplification-specific constant (see Supplemental material 5). As a result, for testing MAKERGAUL C we used the mean values for Factor inhibition separately for every gene as described above. The complete data sets that we have used to calculate these constants are given in Supplemental material 6.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Absolute bias</th>
<th>Linearity</th>
<th>Reproducibility</th>
<th>Resolution</th>
<th>Mean rank</th>
<th>Friedman subset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cy0</td>
<td>2.13 (2)</td>
<td>3.21 (1)</td>
<td>3.73 (2)</td>
<td>3.330 (2)</td>
<td>1.75 (1)</td>
<td>1</td>
</tr>
<tr>
<td>LinRegPCR</td>
<td>6.60 (4)</td>
<td>4.02 (2)</td>
<td>2.63 (1)</td>
<td>2.780 (1)</td>
<td>2.00 (2)</td>
<td>1</td>
</tr>
<tr>
<td>Standart-Cq</td>
<td>2.10 (1)</td>
<td>4.57 (3)</td>
<td>4.30 (3)</td>
<td>4.270 (3)</td>
<td>2.50 (3)</td>
<td>1</td>
</tr>
<tr>
<td>MAKERGAUL_C</td>
<td>6.30 (3)</td>
<td>5.21 (5)</td>
<td>5.83 (6)</td>
<td>5.370 (5)</td>
<td>4.75 (4)</td>
<td>2</td>
</tr>
<tr>
<td>PCR_Miner</td>
<td>8.24 (9)</td>
<td>4.90 (4)</td>
<td>5.13 (4)</td>
<td>4.800 (4)</td>
<td>5.25 (5)</td>
<td>2</td>
</tr>
<tr>
<td>MAK2</td>
<td>7.30 (6)</td>
<td>5.59 (6)</td>
<td>5.70 (5)</td>
<td>5.590 (6)</td>
<td>5.75 (6)</td>
<td>2.5</td>
</tr>
<tr>
<td>MAKERGAUL</td>
<td>6.83 (5)</td>
<td>6.46 (8)</td>
<td>6.14 (7)</td>
<td>6.160 (8)</td>
<td>7.00 (7)</td>
<td>3</td>
</tr>
<tr>
<td>LRE qPCR E100</td>
<td>7.46 (7)</td>
<td>5.76 (7)</td>
<td>6.19 (8)</td>
<td>6.130 (7)</td>
<td>7.25 (8)</td>
<td>3</td>
</tr>
<tr>
<td>5PSM</td>
<td>9.08 (12)</td>
<td>7.84 (9)</td>
<td>7.70 (9)</td>
<td>8.030 (9)</td>
<td>9.75 (9)</td>
<td>4</td>
</tr>
<tr>
<td>DART</td>
<td>8.67 (10)</td>
<td>9.75 (10)</td>
<td>9.78 (10)</td>
<td>10.080 (10)</td>
<td>10.00 (10)</td>
<td>4</td>
</tr>
<tr>
<td>FPLM</td>
<td>7.95 (8)</td>
<td>10.44 (11)</td>
<td>10.73 (11)</td>
<td>10.590 (11)</td>
<td>10.25 (11)</td>
<td>4.5</td>
</tr>
<tr>
<td>LRE qPCR Emax</td>
<td>9.29 (13)</td>
<td>11.08 (12)</td>
<td>11.43 (12)</td>
<td>11.490 (12)</td>
<td>12.25 (12)</td>
<td>5.6</td>
</tr>
<tr>
<td>FPK_PCR</td>
<td>9.06 (11)</td>
<td>12.17 (13)</td>
<td>11.71 (13)</td>
<td>12.300 (13)</td>
<td>12.50 (13)</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 1: Comparative analysis of absolute bias, linearity, reproducibility and resolution of each method.

Please note: The first four columns contain the rank of the method for the performance parameters absolute bias, linearity, reproducibility, and resolution. It ranks the methods’ average performance over all 63 genes. The number in parentheses shows the rank of the methods’ performance per indicator. The fifth column contains the methods’ average rank overall performance parameters. The last column shows the subgroups of methods which have a similar performance considering the four parameters.
Resulting software

The final program is primarily designed for installation as a server-based application. The principal data flow in MAKERGAUL and its individual modules are depicted in Fig. 2. The operator termed “USER” provides relevant information and sends a request to the server harboring the MAKERGAUL program. MAKERGAUL contains individual components that have a modular structure. In the main menu the user classifies his samples and provides information on number of wells and rows that should be quantified from respective plates. Alternatively, preformed data sets that were already created on a real time PCR system can be directly imported. In addition, the user can change data analysis and output formats. A typical example screen that provides a good impression of the program surface is depicted in Fig. 3. When the user has incorporated or imported his data to be analyzed, the programs can be executed and provide quantification data that can be easily exported in formats that allow setting up of comparative expression graphs.

Noteworthy, as an important feature for evolving and improving the software by other researchers, all components are designed and published as open source software. Since there are some free-to-use web-based solutions for other methods (e.g., [42] and [43]), we also want to share the complete scripts (Supplemental material 3) which enable everybody to set up own servers and add new algorithms.

Performance outcome

After developing and evaluating our two algorithms, we next compared them with other established real-time PCR analysis methods (see Model evaluation section and Table 1). In this analysis we could show that the variant MAKERGAUL_C with fixed values for Factor-inhibition has a convincingly better mean ranking than MAKERGAUL and also a better one than the MAK2 algorithm, although this difference is not statistically significant. Based on these and other findings, we recommend determining first the Factor-inhibition by MAKERGAUL in all samples of a particular experiment with similar primers. After that, building
a gene-specific average value for Factor_inhibition that is used for reanalysis with MAKERGAUL_C will lead to better quantification results. As already described above, Factor_inhibition shows no concentration dependency suggesting that there is no need to establish a special calibration setup or dilution series for individual target sequences. We further suggest that a once generated average for Factor_inhibition is suitable for quantification in other experiments that use the same primers resulting in similar amplicons.

The detailed comparison of the different models revealed that three other methods have a significantly better average performance than the MAKERGAUL_C algorithm (see Table 1, Friedman subset 1). Nevertheless, these methods have other limitations. Two of them (Cy0 and Standard-Cq) need preparation of a standard curve for calibration in every real time PCR experiment [30,32]. LinRegPCR as the third superior method requires the determination of the individual PCR efficiency of every sample and calculation of the mean efficiency by averaging at least two samples using the same primer combination (producing the same amplicon) for reaching the shown performance [30,36]. MAKERGAUL and MAKERGAUL_C do not have these requirements on experimental design or individual probes. Therefore, choosing MAKERGAUL/MAKERGAUL_C for data analysis is in our view particularly useful when dealing with large sets of samples resulting in identical amplicons.

A further comparison of the execution times between MAKERGAUL and its relative MAK2 was not possible because the MAK2 algorithm in our implementation is somewhat handicapped by the fact that this algorithm would need to incorporate an additional library to perform these kinds of studies (see Material and methods). Nevertheless, there is a general advantage in MAKERGAUL due to the fact that this algorithm does not require computation of logarithms, which needs iterations of basic arithmetic functions, lookup tables or both [44]. Moreover, the number of required iterations and lookup tables dramatically extends with requested precision. In our case, in which the compared models have to deal with numbers from 1E3 to 1E-14 and further need exact computation even down to this precision, it is hard to establish an implementation of a function that can beat an algorithm that only contains fast basic arithmetic operations.

Conclusion

In summary, MAKERGAUL shows an overall good precision, linearity, and resolution in calculating DNA quantities over a great range of genes and concentrations. Furthermore, the analysis with this algorithm has short execution times. The MAKERGAUL_C implementation complements the method with the possibility to raise the level of performance above MAK2 by using an amplicon-specific constant that can be determined once for all experiments (see Performance outcome section).

However, there are still some important points that need improvement. Like with the MAK2 model, the analysis of reactions performed with competimer representing primers that bind to target sequences but failed to be extended is still not possible with MAKERGAUL. Also improvement of sample quality control and noise reduction of baseline fluorescence are conceivable that we must admit are presently somewhat more perfect in the LinRegPCR implementation [45].

In summary, we hope that our study and the establishment of the open source software MAKERGAUL will improve the performance and accuracy of nucleic acid quantification. In addition, the disclosure of all
source codes provides in our view a solid basis for developing improved software packages in which existing limitations of quantification are eliminated.

Conflict of interest

The authors have nothing to declare.

Acknowledgments

We would like to thank Jan M. Ruijter for kindly providing supplementary and unpublished data of his studies [30] and providing support to it, especially supplying us with the above mentioned Standard-Cq quantification data. Moreover, the authors would like to thank Ryan J. Russell who posted the minGW-w64-toolchain at sourceforge.net [46] and helped the authors in the understanding and configuration of this tool for our purposes. RW is supported by a grant from the German Research Foundation (SFB TRR 57, P13) and a grant from the IZKF Aachen. The sponsors had no influence on study design or report writing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/jclinbiochem.2013.10.017.

References