Modeling qPCR Curves: Fall 2018 Semester Report

Miranda Parrish

BIOL 504

Modeling qPCR Curves: Fall 2018 Semester Report

Constant Multiplier Method

The previous version of the `fpredbFtaqman` function required an estimated kinetic parameter for the primer and another for the probe. The constant multiplier method was developed to eliminate the estimated kinetic parameter for the probe; only the primer kinetic parameter is estimated and the probe kinetic parameter is calculated

$$K_b = cK_p$$

The calculation of the constant c varies depending on which of the two models is used.

Model 1

Model 1 assumes reversible dissociation/re-association reactions between single template strands, A_1 and A_2, and between primer/hydrolysis probe, P, and a single strand of template, A.

$$A_1 + A_2 \underset{k_{d12}}{\overset{k_{a12}}{\rightleftharpoons}} A_1 - A_2$$

$$A + P \underset{k_d}{\overset{k_a}{\rightleftharpoons}} A - P \qquad \text{(Cobbs, 2012, p. 4)}$$

The K primer and probe parameters used in Model 1 have the form

$$K = \frac{k_d}{k_a}$$

Since the equilibrium constant of the reaction is

$$K_{eq} = \frac{k_a}{k_d} = \exp\left(-\frac{\Delta G}{RT}\right) \qquad \text{(Marimuthu \& Chakrabarti, 2014, p. 175104)}$$

where T is the temperature, R is the gas constant, and ΔG is the free energy of hybridization. then the K parameters can be written

$$K = \frac{1}{K_{eq}} = \exp\left(\frac{\Delta G}{RT}\right)$$

A constant multipiler can be found given a known annealing temperature, T, and free energy of hybridization, ΔG.

Nearest Neighbors Method. The nearest neighbors method can be used to estimate the thermodynamic parameters of DNA hybridization from a sequence of base pairs (SantaLucia, 1998; Marimuthu & Chakrabarti, 2014). This method treats the interaction between two complementary strands of DNA as a series of interactions between neighboring base pairs. The predicted free energy parameter is found using

$$\Delta G^{\circ}_{37}(\text{predicted total}) = \Delta G^{\circ}(\text{initiation at terminals}) + \sum \Delta G^{\circ}(\text{interactions})$$

and a table of experimentally determined standard free energy values for the interactions. The following table, adapted from the unified experimental values reported by SantaLucia (1998, p. 1462), was used to calculate the predicted free energy of hybridization from the sequence of a single strand:

	Sequence $(5' - 3')$	$\mathbf{\Delta G^{\circ}_{37}}$ **(kJ/mol)**
	AA, TT	-4.184
	AT	-3.68192
	TA	-2.42672
	CA, TG	-6.0668
	GT, AC	-6.02496
	CT, AG	-5.35552
	GA, TC	-5.4392
	CG	-9.07928
	GC	-9.37216
	GG, CC	-7.69856
Terminal	G, C	4.10032
	A, T	4.30952

Equilibrium Constant at Temperature. The equilibrium constant at 37°C is

$$K_{eq}^{\circ} = \exp\left(-\frac{\Delta G_{37}^{\circ}}{RT^{\circ}}\right)$$

The equilibrium constant at annealing temperature T can be found using the value of K_{eq}° and the van't Hoff equation

$$K_{eq} = K_{eq}^{\circ} \exp\left(-\frac{\Delta H^{\circ}}{R}\left(\frac{1}{T} - \frac{1}{T^{\circ}}\right)\right)$$

where T° is 310.15 K(37°C) and ΔH° is the enthalpy for the sequence at 37°C. Enthalpy values for nearest neighbor pairs at this temperature are also included in the paper by SantaLucia (1998, p. 1462). So the value of parameter K is

$$K = \frac{1}{K_{eq}} = \left(K_{eq}^{\circ}\right)^{-1} \exp\left(\frac{\Delta H^{\circ}}{R}\left(\frac{1}{T} - \frac{1}{T^{\circ}}\right)\right)$$

Constant Multiplier. The K parameters for the probe and primer are denoted K_b and K_p, respectively.

$$K_b = cK_p$$

$$K_{eq,b}^{-1} = cK_{eq,p}^{-1}$$

$$\left(K_{eq,b}^{\circ}\right)^{-1} \exp\left(\frac{\Delta H_b^{\circ}}{R}\left(\frac{1}{T} - \frac{1}{T^{\circ}}\right)\right) = c\left(K_{eq,p}^{\circ}\right)^{-1} \exp\left(\frac{\Delta H_p^{\circ}}{R}\left(\frac{1}{T} - \frac{1}{T^{\circ}}\right)\right)$$

$$c = \frac{K_{eq,p}^{\circ}}{K_{eq,b}^{\circ}} \exp\left(\frac{\Delta H_b^{\circ} - \Delta H_p^{\circ}}{R}\left(\frac{1}{T} - \frac{1}{T^{\circ}}\right)\right)$$

Model 2

Model 2 assumes irrreversible dissociation/re-association reactions between single template strands, A_1 and A_2, and between primer/hydrolysis probe, P, and a single strand of template, A.

$$A_1 + A_2 \xrightarrow{k_{a12}} A_1 - A_2$$

$$A + P \xrightarrow{k_a} A - P \qquad \text{(Cobbs, 2012, p. 4)}$$

The K primer and probe parameters used in Model 2 have the form

$$K = \frac{k_a}{k_{a12} - k_a}$$

Proportionality of Parameters. The forward rate constant for the association of two strands of DNA is proportional to the square root of the length in base pairs of the shorter strand.

$k \sim \sqrt{N}$ (Craig, Crothers, & Doty, 1971; Mehra & Hu, 2005; Wetmur & Davidson, 1968)

In their model, Mehra and Hu (2005, p. 851) used a value of $k_a \sim \sqrt{N} k_{a12}$. The constant for Model 2 is currently based on an assumption that the forward rate constant k_{a12} is

$$k_{a12} = \frac{k_a}{\sqrt{N}}$$

Constant Multiplier. The K parameters for the probe and primer are denoted K_b and K_p, respectively. The lengths of the primer and probe are N_p and N_b.

$$\begin{aligned} K_b &= cK_p \\ \frac{k_b}{k_{a12} - k_b} &= c\,\frac{k_p}{k_{a12} - k_p} \\ \frac{k_b}{k_b N_b^{-0.5} - k_b} &= c\,\frac{k_p}{k_p N_p^{-0.5} - k_p} \\ \frac{1}{N_b^{-0.5} - 1} &= c\,\frac{1}{N_p^{-0.5} - 1} \\ c &= \frac{N_p^{-0.5} - 1}{N_b^{-0.5} - 1} \end{aligned}$$

R Code

fpredbFtaqman(...)

The `fpredbFtaqman(...)` was modified to accept a `constant` parameter rather than an estimated `Kprobe` parameter. The value of `Kprobe` was set to `constant*Kprimer`.

Old Version.

```
fpredbFtaqman <- function(S00, Kf, Kprimer, Kprobe) { ... }
```

New (Constant Multiplier) Version.

```
fpredbFtaqman <- function(S00, Kf, Kprimer, constant = 1, modelno = 2) {
    Kprobe <- (constant * Kprimer)
    . . .}
```

nnTab.txt

The file `nnTab.txt` is the table of nearest neighbor sequences and corresponding ΔG°_{37} values used for predicting ΔG°_{37} from a sequence. It will likely include a `'deltaH'` column and corresponding ΔH° values in the future.

```
'Sequence'  'deltaG'
'AA'    -4.184
'AT'    -3.68192
'TA'    -2.42672
'CA'    -6.0668
'GT'    -6.02496
'CT'    -5.35552
'GA'    -5.4392
'CG'    -9.07928
'GC'    -9.37216
'GG'    -7.69856
'TT'    -4.184
'TG'    -6.0668
'AC'    -6.02496
'AG'    -5.35552
'TC'    -5.4392
'CC'    -7.69856
'G' 4.10032
'C' 4.10032
'A' 4.30952
'T' 4.30952
```

getDeltaG(...)

The `getDeltaG` function uses the nearest neighbors method to calculate a predicted free energy of hybridization at 37°C for a given sequence `seq`. The function creates a list,

`neighbors`, of substrings taken from the `seq` parameter: the first base, last base, and every pair of neighboring bases. The `nntab` object is the table read from the `nnTab.txt` file. The object `ind` lists the row index of the matching sequence in the `'Sequence'` column of `nntab` for each element in `neighbors`. This list of indices is used to retrieve the corresponding `'deltaG'` column of `nntab` . The sum of these values is then returned.

```
getDeltaG <- function(seq) {
    neighbors <- list()
    ind <- list()
    # Get the terminal bases and each pair of "neighbors" from the sequence
    neighbors <- c(substr(seq, 1, 1),
        substr(seq, nchar(seq), nchar(seq)),
        mapply(substr, seq, 1:(nchar(seq)-1), 2:nchar(seq), USE.NAMES=FALSE))
    # For each neighbor, get the index of where its sequence appears in the
       Sequence column of the nearest neighbors table
    for (i in 1:length(neighbors)) {
        ind <- append(ind, which(neighbors[i] == nntab$Sequence,  arr.ind =
           TRUE))
        }
    ind <- unlist(ind)
    # Get and sum the deltaG values from the deltaG column which correspond
       to the neighbors
    deltaG <- sum(nntab$deltaG[ind])
    return(deltaG)
    }
```

getConstant(...)

The `getConstant` function returns a constant multiplier, `c`, for the given probe and primer sequences according to the specified model. While the function requires only probe

and primer sequences to determine a constant for Model 2, determining a constant under Model 1 assumptions will require an additional parameter, `ctemp`, which is the annealing temperature for the reaction in degrees Celsius. The conversion of K°_{eq} to K_{eq} for a given temperature has not yet been implemented and will be required to find a constant for Model 1.

```
getConstant <- function(primerseq, probeseq, model = 2, ctemp = 0){
    if (model == 1) # UNDER CONSTRUCTION
    if (model == 2) {
        Nb <- nchar(probeseq)
        Np <- nchar(primerseq)
        c <- ((1 / sqrt(Np)) - 1) / ((1 / sqrt(Nb)) - 1)
        }
    return(c)
    }
```

Cyp1B1

The Cyp1B1 data set from Smith, Miller, Kohn, Walker, and Portier (2007) was used to test the constant multiplier method. Three replicate PCRs were performed on solutions with 10^3, 10^4, 10^5, and 10^6 initial target DNA molecules for a total of 12 PCRs on the Cyp1B1 gene. The sequence of the forward primer was "GCCAGCCAGGACACCCT", the sequence of the TaqMan probe was "CGCTTGCAGTGGCTGCTCCTCCT", and the annealing temperature was 60°C (Smith et al., 2007).

Model 1 Constant

For the forward primer sequence used for the Cyp1B1 dataset, the nearest neighbors method estimates ΔG° from the interactions

$$\begin{array}{ll} \mathbf{5'} & \mathrm{G}\overset{\downarrow}{-}\mathrm{C}\overset{\downarrow}{-}\mathrm{C}\overset{\downarrow}{-}\mathrm{A}\overset{\downarrow}{-}\mathrm{G}\overset{\downarrow}{-}\mathrm{C}\overset{\downarrow}{-}\mathrm{C}\overset{\downarrow}{-}\mathrm{A}\overset{\downarrow}{-}\mathrm{G}\overset{\downarrow}{-}\mathrm{G}\overset{\downarrow}{-}\mathrm{A}\overset{\downarrow}{-}\mathrm{C}\overset{\downarrow}{-}\mathrm{A}\overset{\downarrow}{-}\mathrm{C}\overset{\downarrow}{-}\mathrm{C}\overset{\downarrow}{-}\mathrm{C}\overset{\downarrow}{-}\mathrm{T} & \mathbf{3'} \\ \mathbf{3'} & \mathrm{C}-\mathrm{G}-\mathrm{G}-\mathrm{T}-\mathrm{C}-\mathrm{G}-\mathrm{G}-\mathrm{T}-\mathrm{C}-\mathrm{C}-\mathrm{T}-\mathrm{G}-\mathrm{T}-\mathrm{G}-\mathrm{G}-\mathrm{G}-\mathrm{A} & \mathbf{5'} \end{array}$$

and equation

$$\begin{aligned} \Delta G^\circ_{\text{primer}} &= \Delta G^\circ(\text{init.}) + \sum \Delta G^\circ(\text{interactions}) \\ &= \Delta G^\circ(\text{G/C init.}) + \Delta G^\circ(\text{GC/CG}) + \Delta G^\circ(\text{CC/GG}) + \Delta G^\circ(\text{CA/GT}) \\ &\quad + \ \ldots \\ &\quad + \Delta G^\circ(\text{CC/GG}) + \Delta G^\circ(\text{CC/GG}) + \Delta G^\circ(\text{CT/GA}) + \Delta G^\circ(\text{A/T init.}) \\ &= 4.10032 - 9.37216 - 7.69856 - 6.0668 + \ \ldots \\ &\quad - 7.69856 - 7.69856 - 5.35552 + 4.30952 \\ &= -100.5834 \end{aligned}$$

For the probe, $\Delta G^\circ_{\text{probe}} = -138.7414$. The calculation of the constant for Model 1 has not yet been implemented.

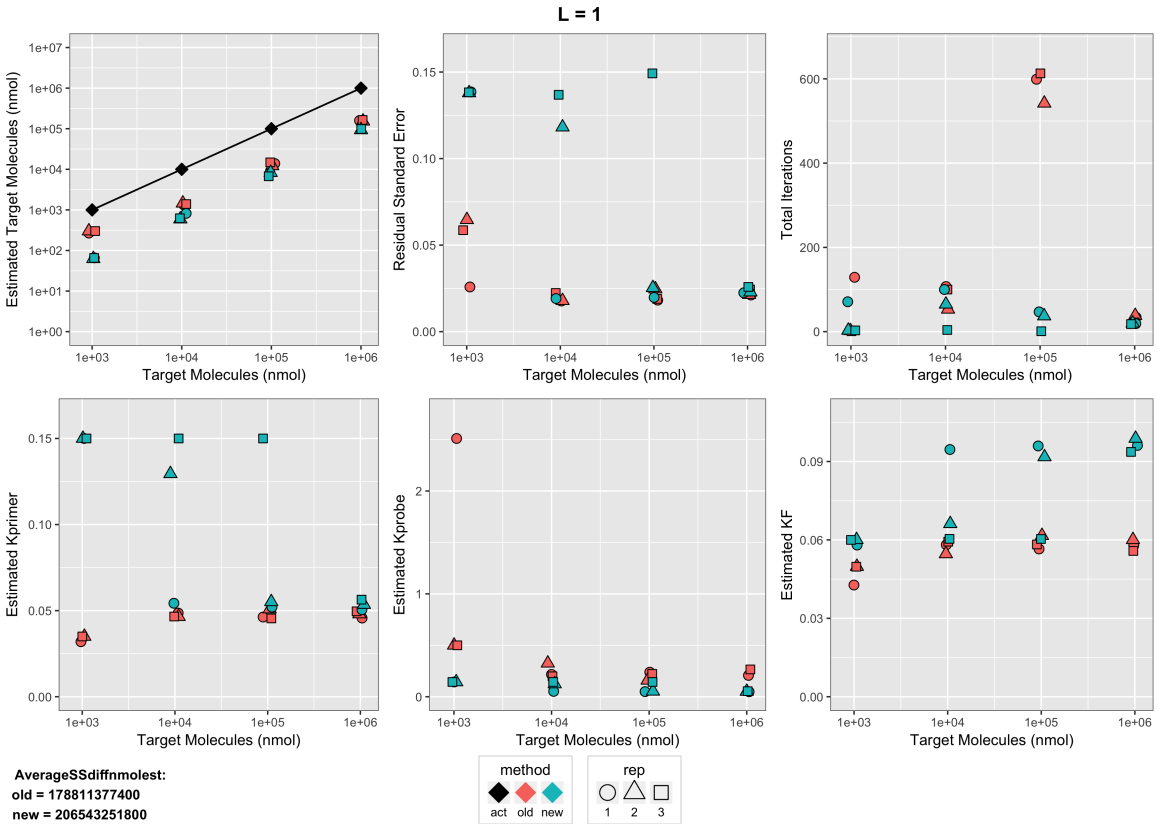
Model 2 Constant

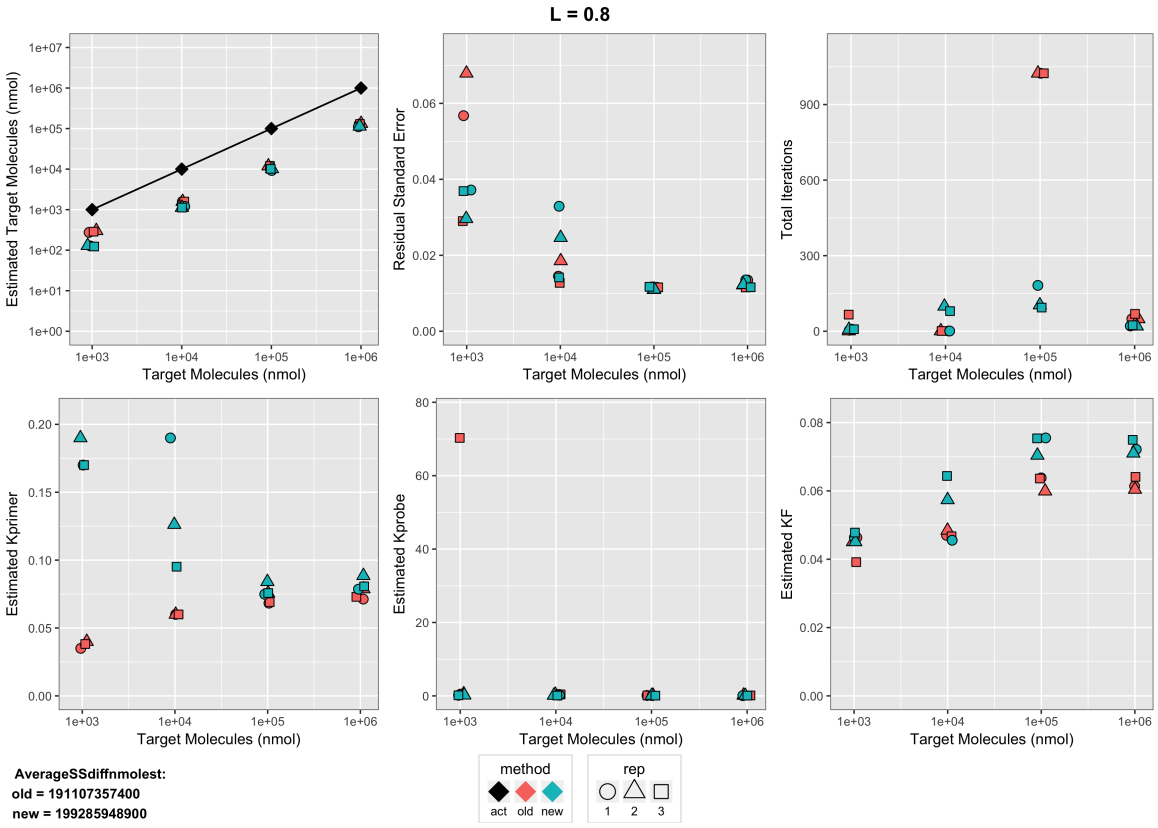
The lengths of the primer and probe sequences are $N_p = 17$ and $N_b = 23$, so the constant is

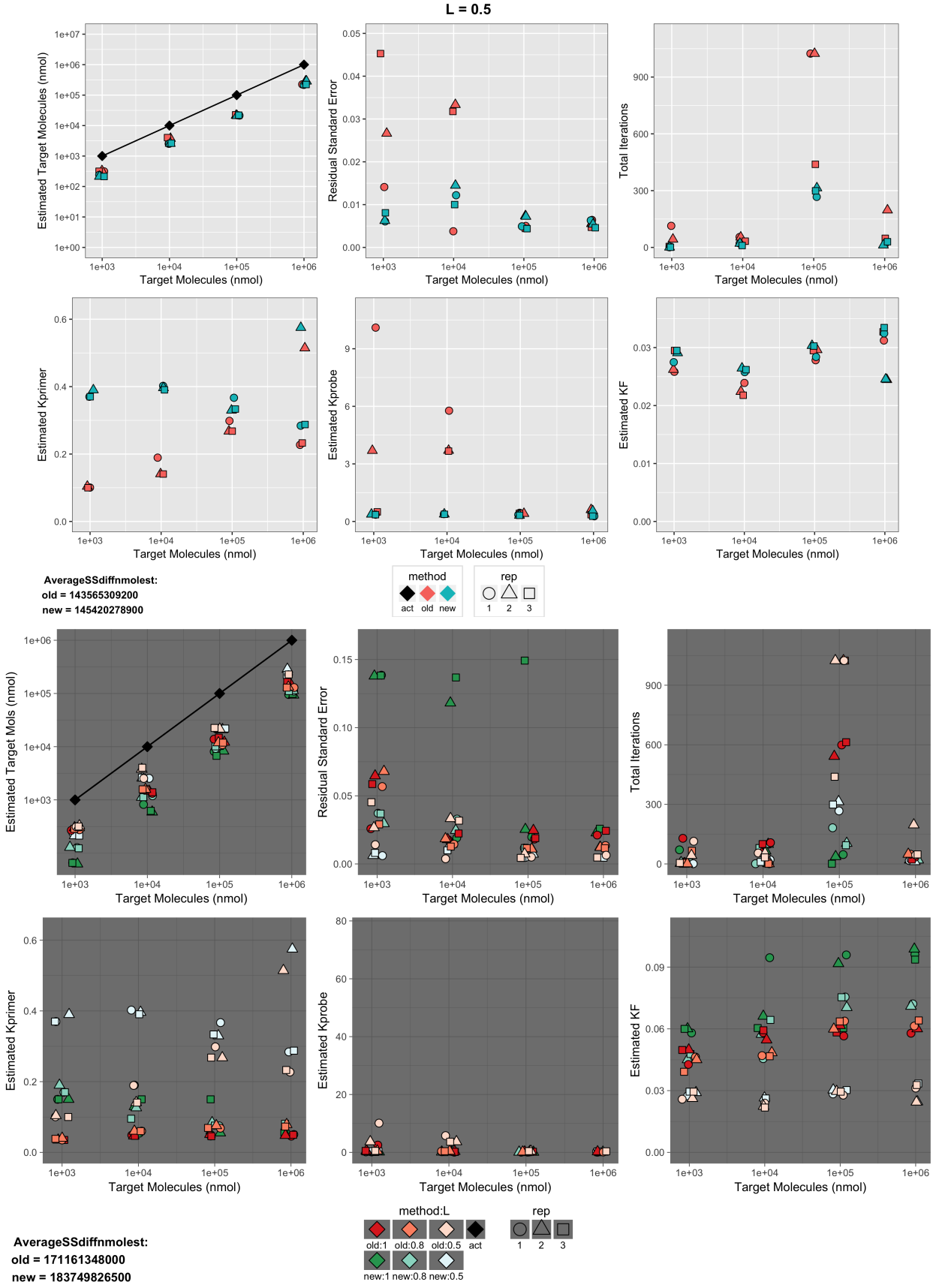
$$c = \frac{N_p^{-0.5} - 1}{N_b^{-0.5} - 1} = \frac{17^{-0.5} - 1}{23^{-0.5} - 1} = 0.957016$$

Results

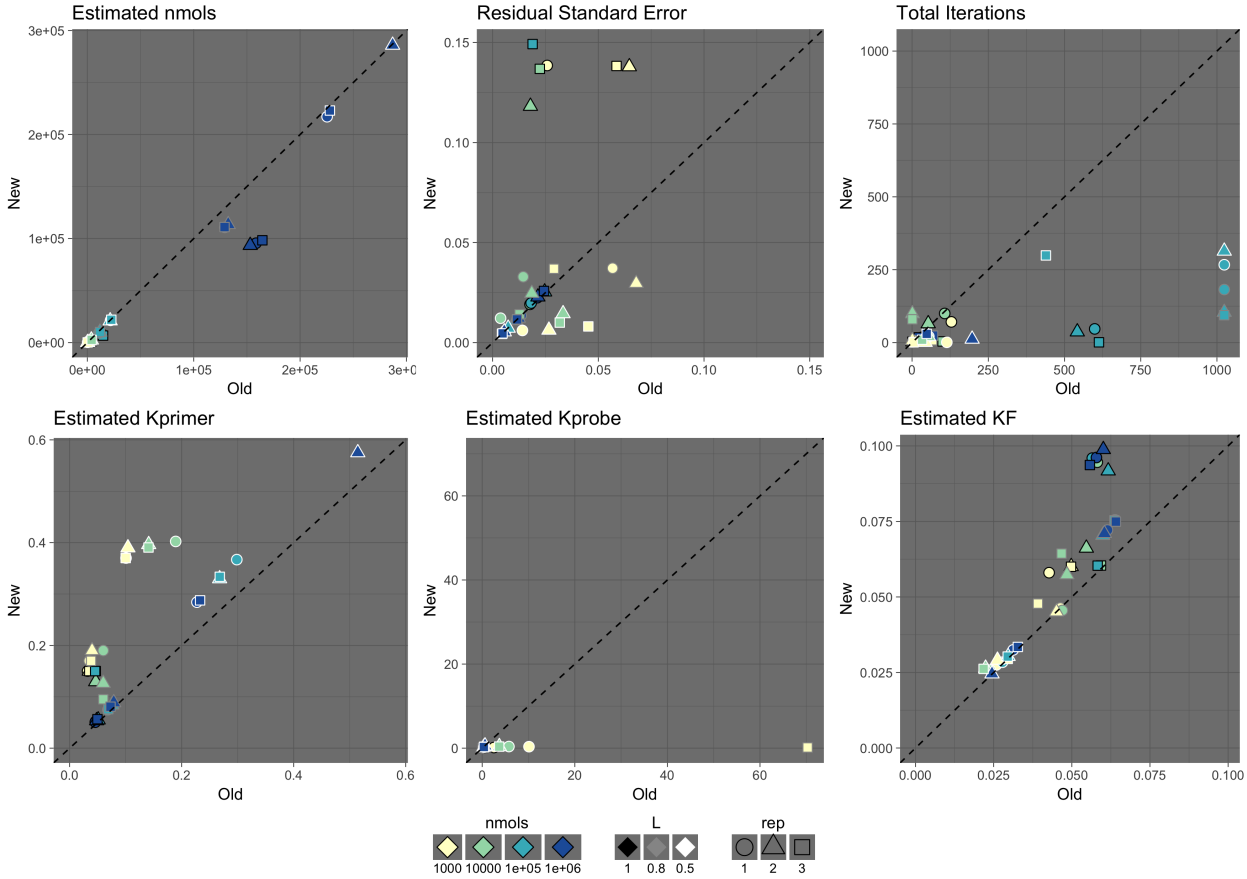
The old method and new, constant multiplier method were applied to the Cyp1B1 dataset for three different L cut-off values: 1, 0.8, and 0.5. These L values are used to restrict the model fit to only part of the curve. The entire curve is fit at $L = 1$; decreasing the value of L further restricts the model fit.


L = 1
Estimated Target Molecules (nmol)
Residual Standard Error
Total Iterations
Estimated Kprimer
Estimated Kprobe
Estimated KF
Target Molecules (nmol)
AverageSSdiffnmolest:
old = 178811377400
new = 206543251800
method
act old new
rep
1 2 3


L = 0.8
Estimated Target Molecules (nmol)
Residual Standard Error
Total Iterations
Estimated Kprimer
Estimated Kprobe
Estimated KF
Target Molecules (nmol)
AverageSSdiffnmolest:
old = 191107357400
new = 199285948900
method
act old new
rep
1 2 3


L = 0.5
Estimated Target Molecules (nmol)
Residual Standard Error
Total Iterations
Estimated Kprimer
Estimated Kprobe
Estimated KF
Target Molecules (nmol)
AverageSSdiffnmolest:
old = 143565309200
new = 145420278900
method
act old new
rep
1 2 3
Estimated Target Mols (nmol)
AverageSSdiffnmolest:
old = 171161348000
new = 183749826500
method:L
old:1 old:0.8 old:0.5 act
new:1 new:0.8 new:0.5

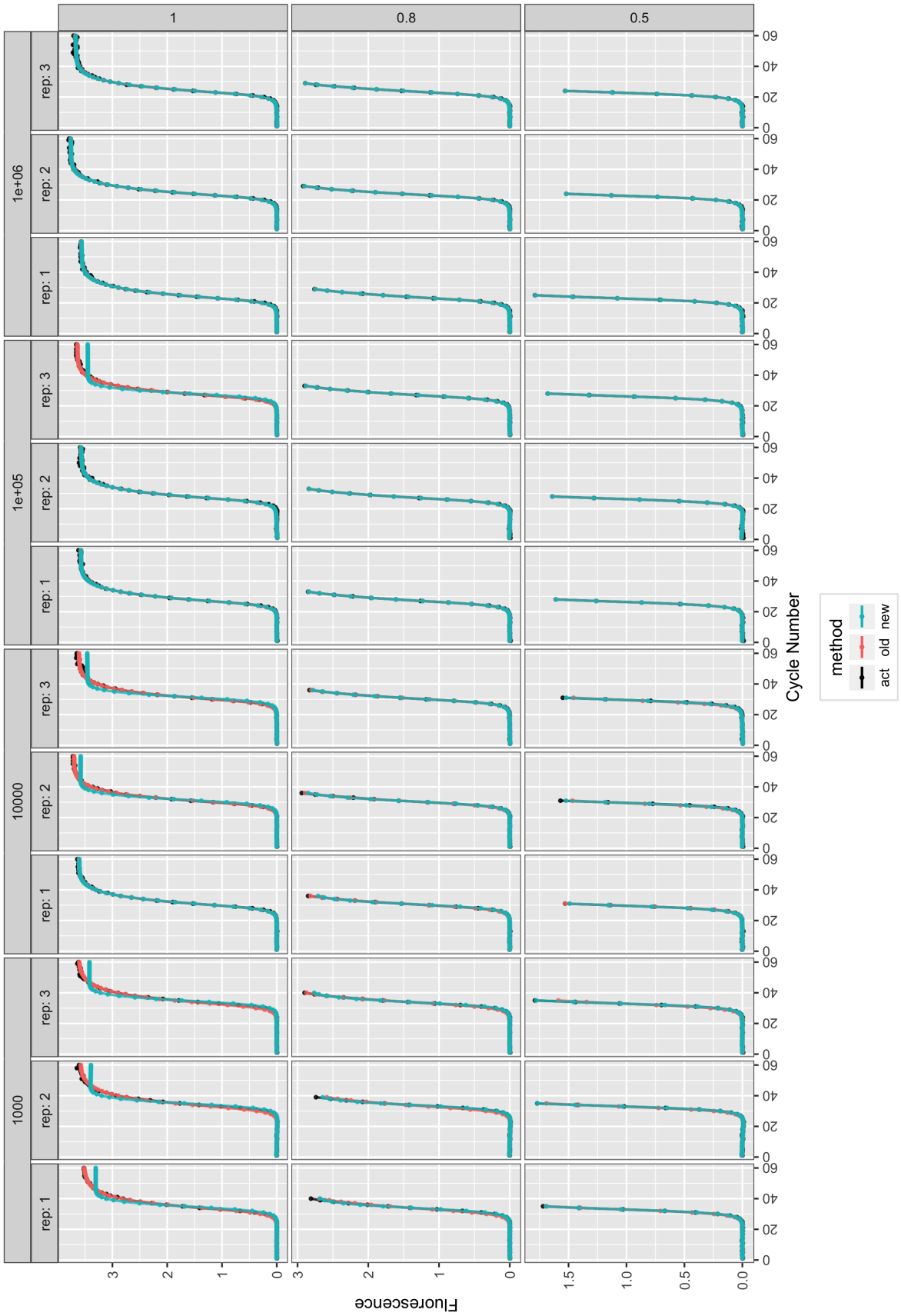
Method Comparison

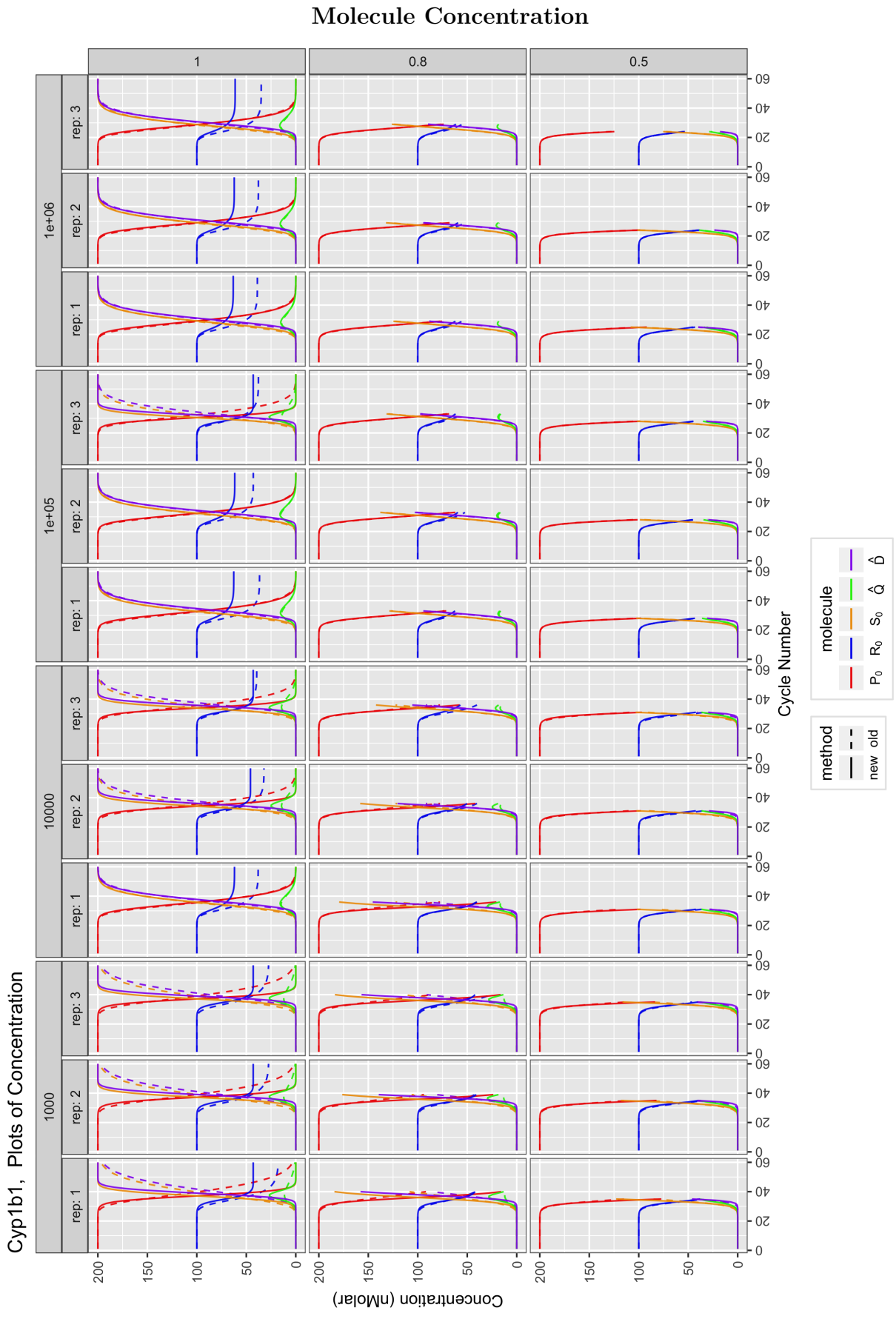


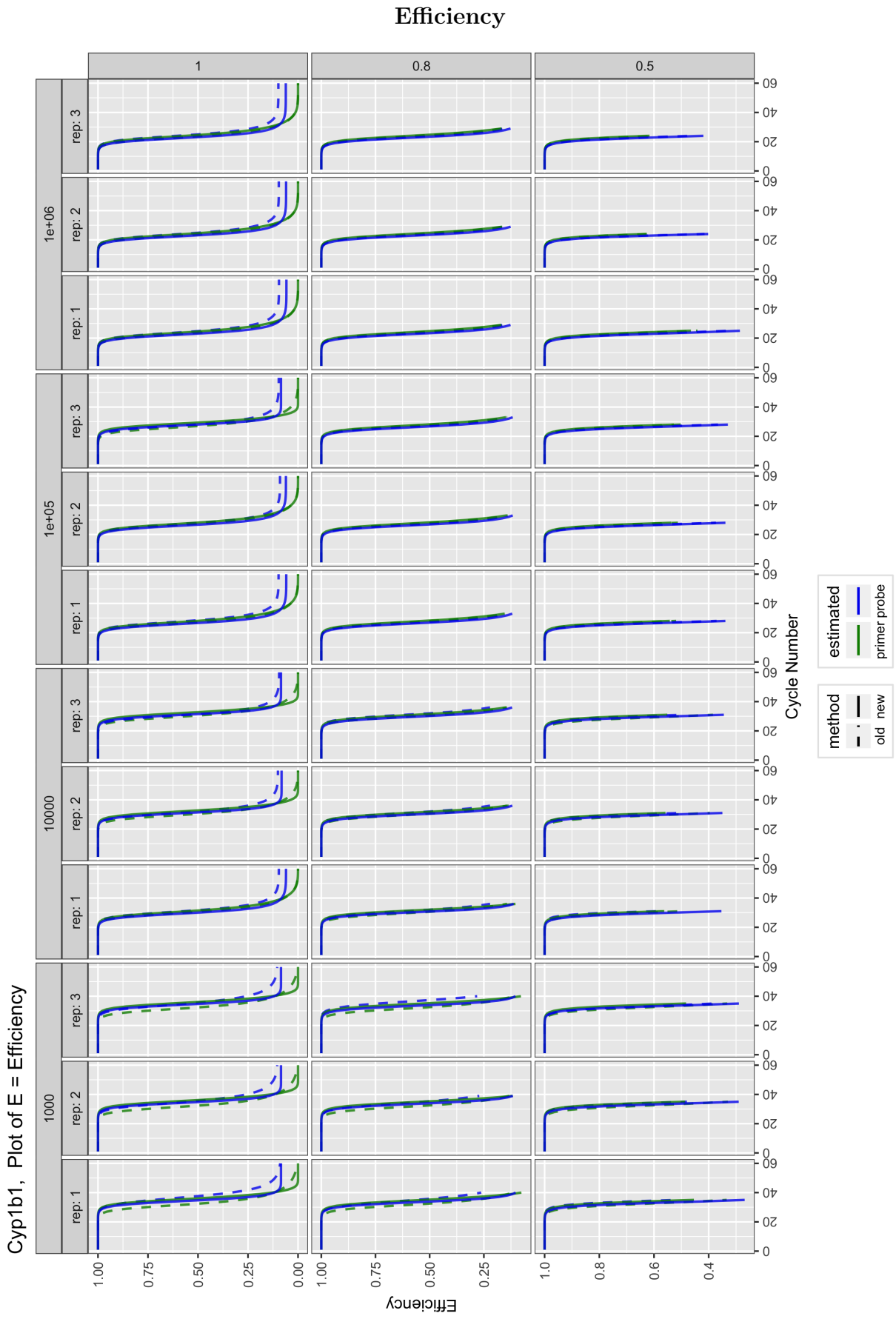
Avg Sum of Squares Difference Between NMol Est. and Actual Value

method	L Value			
	1	0.8	0.5	avg
new	206, 543, 251, 762	199, 285, 948, 916	145, 420, 278, 910	183, 749, 826, 530
old	178, 811, 377, 378	191, 107, 357, 400	143, 565, 309, 185	171, 161, 347, 988

Fit to qPCR Curve




Molecule Concentration
Cyp1b1, Plots of Concentration
1
0.8
0.5
1e+06
1e+05
10000
1000
rep: 1
rep: 2
rep: 3
Concentration (nMolar)
Cycle Number
molecule
P_0 R_0 S_0 $\hat{Q}$ $\hat{D}$
method
new old


Efficiency
1
0.8
0.5
1e+06
1e+05
10000
1000
rep: 1
rep: 2
rep: 3
Cyp1b1, Plot of E = Efficiency
Cycle Number
Efficiency
estimated
primer probe
method
old
new

Conclusions

The constant multiplier method performs more poorly than the old method when fitting the entire curve; it gives both a worse estimate of initial target molecules and worse fit to the qPCR curve.

However, when the dataset is truncated and the later cycles of the qPCR curve removed from model fitting, the fit of the constant multiplier method to the curve seems about the same as or better (especially for the lower initial target concentrations) than that of the old method. In addition, the estimate of initial target concentration is improved by truncating the data and is more similar to that of the old method.

Given the difference in performance based on on the amount of the curve fit and from looking at the fluorescence curves, the method appears less accurate when fitting cycles occuring after the exponential phase. Including processivity in the model may improve the fit to the curve and estimate of initial target concentration.

References

Cobbs, G. (2012, Aug). Stepwise kinetic equilibrium models of quantitative polymerase chain reaction. *BMC Bioinformatics*, *13*, 203. Retrieved from https://doi.org/10.1186/1471-2105-13-203 doi: 10.1186/1471-2105-13-203

Craig, M. E., Crothers, D. M., & Doty, P. (1971, Dec). Relaxation kinetics of dimer formation by self complementary oligonucleotides. *Journal of Molecular Biology*, *62*(2), 383-401. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/0022283671904347 doi: 10.1016/0022-2836(71)90434-7

Marimuthu, K., & Chakrabarti, R. (2014, May). Sequence-dependent theory of oligonucleotide hybridization kinetics. *The Journal of Chemical Physics*, *140*(17), 175104. Retrieved from https://aip.scitation.org/doi/10.1063/1.4873585 doi: 10.1063/1.4873585

Mehra, S., & Hu, W.-S. (2005, Sep). A kinetic model of quantitative real-time polymerase chain reaction. *Biotechnology and Bioengineering*, *91*(7), 848-860. Retrieved from http://onlinelibrary.wiley.com/doi/abs/10.1002/bit.20555 doi: 10.1002/bit.20555

SantaLucia, J. (1998, Feb). A unified view of polymer, dumbbell, and oligonucleotide dna nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(4), 1460-1465. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC19045/

Smith, M. V., Miller, C. R., Kohn, M., Walker, N. J., & Portier, C. J. (2007, Oct). Absolute estimation of initial concentrations of amplicon in a real-time rt-pcr process. *BMC Bioinformatics*, *8*(1), 409. Retrieved from https://doi.org/10.1186/1471-2105-8-409 doi: 10.1186/1471-2105-8-409

Wetmur, J. G., & Davidson, N. (1968, Feb). Kinetics of renaturation of dna. *Journal of Molecular Biology*, *31*(3), 349-370. Retrieved from

http://linkinghub.elsevier.com/retrieve/pii/0022283668904142 doi: 10.1016/0022-2836(68)90414-2