Modeling qPCR Curves: Spring 2019 Semester Report

Miranda Parrish

BIOL 405

Modeling qPCR Curves: Spring 2019 Semester Report

Model 1 Changes

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.

$$A_1 + A_2 \underset{k_{diss}}{\overset{k_{ass}}{\rightleftharpoons}} A_1 - A_2$$

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

The K_S parameter for the primer and probe used in Model 1 have the form

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

and the K_D parameter has the form

$$K_D = \frac{k_{diss}}{k_{ass}}$$

The model must estimate numerous parameters for every replicate in a dataset. Determining starting values for every combination of parameter values is a lengthy process. The time required to estimate these parameters is only slightly diminished by only estimating K_S for the primer and assuming K_S for the probe is equal.

The goal of this semester was to reduce the number of parameters estimated by the model and to determine a relationship between the primer and probe K_S parameters.

Parameter Reduction: K_D

A literature search was conducted in order to find a theoretical value of K_D that could be used in lieu of estimating the parameter. The following values

$k_{ass} \approx 1 \times 10^6 \ \ M^{-1}\ s^{-1}$ (Craig et al., 1971; Gotoh, 1995; Jensen et al., 1997)

$k_{diss} \approx 1 \times 10^{-4} \ \ s^{-1}$ (Gotoh, 1995; Jensen et al., 1997)

were used to estimate

$$\mathrm{K}_D = 0.1 \text{ nM}$$

Constant Multiplier Relationship: $\mathbf{K}_{S,\,probe} = c * \mathbf{K}_{S,\,primer}$

The old method for Model 1 requires the base sequences of the primer and probe as well as the annealing temperature. The new method for Model 1 is similar to the old method for Model 2 in that both only require the lengths of the primer and probe.

Old Method (Fall 2018). The older method for determining the value of the constant utilizes the Nearest Neighbors Method of estimating the thermodynamic parameters of DNA hybridization from a sequence of base pairs (SantaLucia, 1998), as discussed by Marimuthu and Chakrabarti (2014).

The free energy and enthalpy for the primer and probe at standard temperature is determined from their base sequences using the values reported by SantaLucia (1998, p. 1462). The standard free energy and enthalpy is then used to estimate the equilibrium constant at standard temperature and the van't Hoff equation used to dotermine the equilibrium constant at annealing temperature. The constant for this method is given by

$$c = \frac{K^\circ_{eq,p}}{K^\circ_{eq,b}} \exp\left(\frac{\Delta H^\circ_b - \Delta H^\circ_p}{R}\left(\frac{1}{T} - \frac{1}{T^\circ}\right)\right)$$

where $T^\circ = 310.15$ K(37°C) is the standard temperature, T is the annealing temperature in Kelvin, $K^\circ_{eq,p}$ and ΔH°_p are the equilibrium constant and enthalpy at standard temperature for the primer, and $K^\circ_{eq,b}$ and ΔH°_b are the equilibrium constant and enthalpy at standard temperature for the probe.

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.

$$\mathrm{k} \sim \sqrt{N} \qquad \text{(Craig et al., 1971; Mehra \& Hu, 2005; Wetmur \& Davidson, 1968)}$$

In their model, Mehra and Hu (2005, p. 851) used a value of $k_a \sim k_{ass} * \sqrt{N}$. The new method for the Model 1 constant is based on an assumption that the forward rate constant k_a for the primer and probe is

$$k_a = k_{ass} * \sqrt{N}$$

and that the dissociation constant for the primer and probe is

$$k_d = k_{diss}$$ (Mehra & Hu, 2005, p. 851)

New Method (Current). The new method for determining the value of the constant uses a relationship that follows from the assumptions in the previous section:

$$K_S = \frac{k_{diss}}{k_{ass} * \sqrt{N}}$$

where k_{ass} and k_{diss} are the association and dissociation rate constants for the double stranded DNA template and N is the length of the shorter primer/probe sequence.

The K_S 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.

$$\begin{aligned} K_b &= c * K_p \\ \frac{k_{diss}}{k_{ass} * \sqrt{N_b}} &= c * \frac{k_{diss}}{k_{ass} * \sqrt{N_p}} \\ \sqrt{N_b} &= c * \sqrt{N_p} \\ c &= \sqrt{\frac{N_b}{N_p}} \end{aligned}$$

Code Versions

The following table summarizes the different versions of the code which are compared in the results section.

Version	Estimated Parameters	Given Parameters	K_S Relationship
v0	$K_{S,\,primer}$, K_D	none	$K_{S,\,probe} = K_{S,\,primer}$
v1	$K_{S,\,primer}$, K_D	none	$K_{S,\,probe} = c * K_{S,\,primer}$
v2	$K_{S,\,primer}$	$K_D = 0.1$	$K_{S,\,probe} = c * K_{S,\,primer}$

Version 0 is the unmodified version of the code. The values of both $K_{S,\,primer}$ and K_D are estimated by the code and $K_{S,\,probe}$ is assumed to be equal to $K_{S,\,primer}$.

Version 1 was modified only to include a constant multiplier relationship between $K_{S,\,primer}$ and $K_{S,\,probe}$. The values of both $K_{S,\,primer}$ and K_D are estimated by the code and $K_{S,\,probe}$ is assumed to be equal to $c * K_{S,\,primer}$.

Version 2 was modified so that K_D is given, only the value of $K_{S,\,primer}$ is estimated, and there is a constant multiplier relationship between $K_{S,\,primer}$ and $K_{S,\,probe}$. This version assumes $K_D = 0.1$ and that $K_{S,\,probe}$ is equal to $c * K_{S,\,primer}$.

R Code

getOldConstant(...)

The `getOldConstant` function returns the constant multiplier, `c`, determined by the new method. Given `seqA`, `seqB`, and annealing temperature `tempC` in Celsius, the function returns a constant for the relationship $K_B = c * K_A$. The function uses a table, `nnTab.txt`, of nearest neighbor sequences and corresponding standard free energy and enthalpy values to determine the constant.

```
getOldConstant <- function(seqA, seqB, tempC=NULL]){
    if( is.null(tempC) ){
        stop("No temperature given. Must define temperature when using Model
           1. Input annealing temperature in Celsius.")
        }
    R <- gasConstant
    temp0 <- 310.10 # in K, 37 C
    temp <- tempC + 273.15  # in K
    indA <- getIndex(seqA)  # get indices for nearest neighbors table
    indB <- getIndex(seqB)
    deltaG0A <- sum(nntab$deltaG0[indA])
    deltaG0B <- sum(nntab$deltaG0[indB])
```

```
    Keq0A <- exp(-deltaG0A / (R * temp0))
    Keq0B <- exp(-deltaG0B / (R * temp0))
    deltaH0A <- sum(nntab$deltaH0[indA])
    deltaH0B <- sum(nntab$deltaH0[indB])
    KeqA <- Keq0A * exp( -(deltaH0A / R) * ( (1/temp) - (1/temp0) ) )
    KeqB <- Keq0B * exp( -(deltaH0B / R) * ( (1/temp) - (1/temp0) ) )
    KA <- (1 / KeqA)
    KB <- (1 / KeqB)
    c <- (KB / KA)
    return(c)
    }
```

getNewConstant(...)

The `getNewConstant` function returns the constant multiplier, `c`, determined by the new method. Given `seqA` and `seqB`, the function returns a constant for the relationship $\mathrm{K}_B = c * \mathrm{K}_A$. The function currently requires the sequences of the primer and probe, but it can easily be modified to return a constant from only their lengths.

```
getNewConstant <- function(seqA, seqB) {
    nA <- nchar(seqA)
    nB <- nchar(seqB)
    c <- sqrt(nB / nA)
    return(c)
    }
```

fpredbFtaqmanmod1(...)

The `fpredbFtaqmanmod1` function was modified to allow for testing of the different versions of the code.

Old.

```
fpredbFtaqmanmod1<-function(S00, Kf, KD, KprimerS) { ... }
```

New.

```
fpredbFtaqmanmod1<-function(S00, Kf, KD, KprimerS, version = versionGlobal) {
    ...
    if(version == 0) {
        KprobeS <- KprimerS                # Assume KprobeS = KprimerS
        } else
    if(version == 1 || version == 2 || version == 3) {
        KprobeS <- (constant * KprimerS)    # Assume KprobeS = c * KprimerS
        }
    ... }
```

Other Modifications

The `createstartvaluesmod1` and `nmolplotgenmod1` functions were both heavily modified to allow the code to run differently based on the specified version.

Error Handling. The `nlsLM` function in the results generating file is particularly prone to errors. This section of the results generating file was wrapped in a separate error-handling function that utilized a `tryCatch` function to handle any errors. Each version/truncation run has an error log. If an error is caught the rep which caused it is removed from the results and the error is added to the error log. If any error is caught during the run, the error log for the version/truncation run is exported to a text file at the end of the run. This allowed for numerous runs on various versions and with different truncations of the dataset without stopping for `nlsLM` errors.

Cyp1B1

Model 1 was tested with the same data set—Cyp1B1 (Smith, Miller, Kohn, Walker, & Portier, 2007)—that was used for testing Model 2 during the Fall 2018 semester. 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).

DNA Template/Target DNA

The Smith et al. (2007) paper indicated that the Cyp1B1 template was a section of cDNA that had been reverse-transcribed from an mRNA sequence. Results from a Primer-BLAST (Ye et al., 2012) query using the the complete coding sequence of the Cyp1B1 gene (Sutter et al., 1994) and the forward and reverse primer sequences specified in the Smith et al. (2007) paper indicated that the product was 97 base pairs long.

Constant Multipliers for Cyp1B1

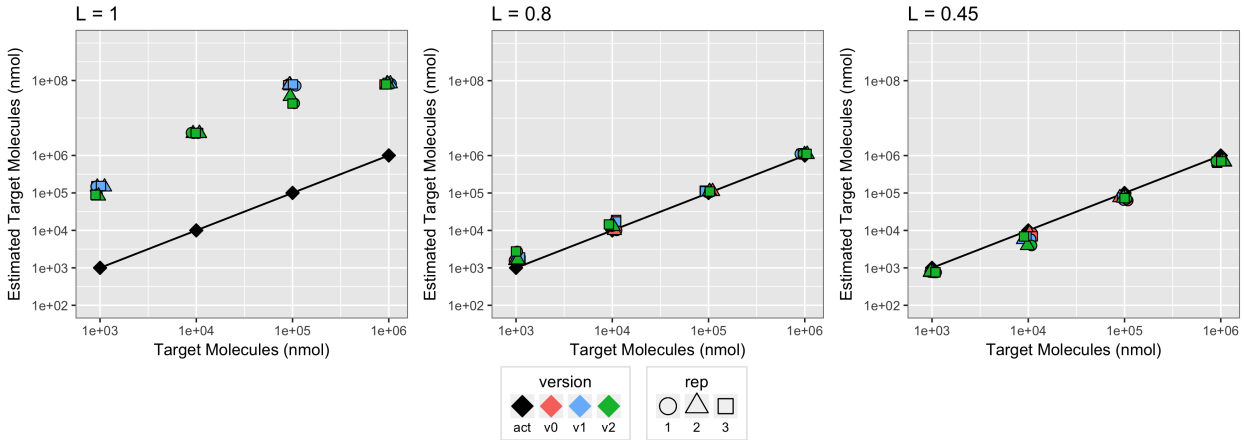
The value of the constant multiplier for the Cyp1B1 dataset was determined using both the old and new method. The old method yielded a constant multiplier of 0.000157014, while new method yielded a constant multiplier of 1.16316.

Results

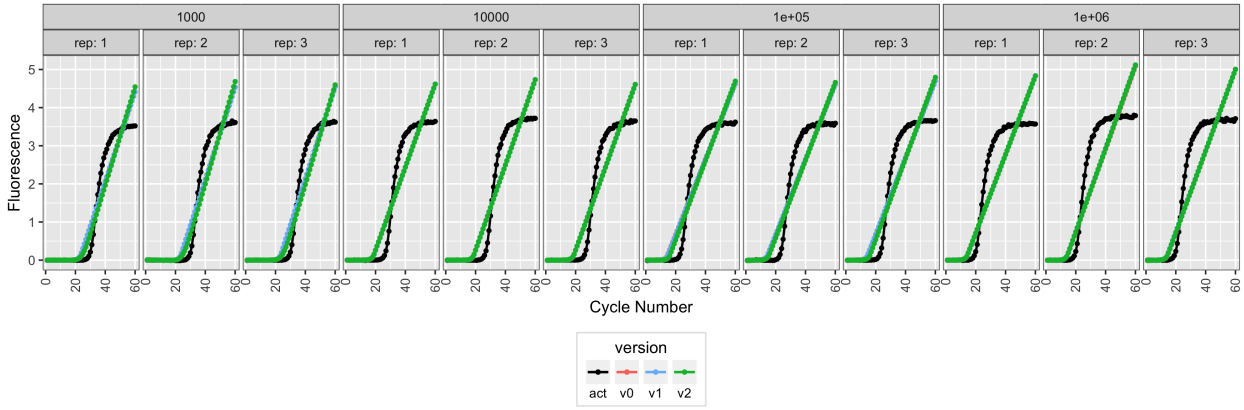
The two constant multiplier methods and various truncations of the data set were tested. Due to poor performance for all versions, the old constant multiplier method and the use of the full data set were not considered when comparing the versions.

Data Truncation

Three different truncations of the Cyp1B1 dataset were used. The L cut-off values were: 1, 0.8, and 0.45. 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. However, all versions of the code performed very poorly when using the the full data set. The estimates for the un-truncated data were much worse than estimates using the truncated data



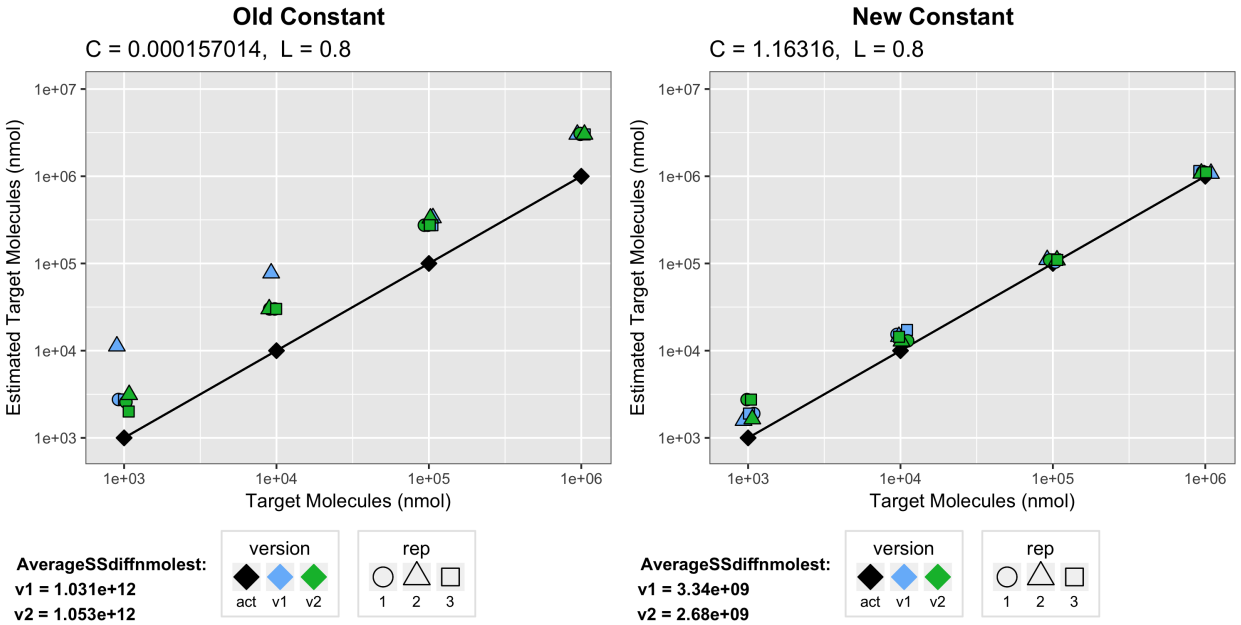
and the fit of the fluorescence curves to the full data set was also very poor:



Given the poor performance of the code on un-truncated data, the performance of the different versions of the code was evaluated using only truncated values ($L = 0.8$, $L = 0.45$).

Constant Multiplier Method Comparison

The performance of the two constant multiplier methods was evaluated using the same truncated ($L = 0.8$) data. Only the results for versions which use a constant multiplier relationship are shown.

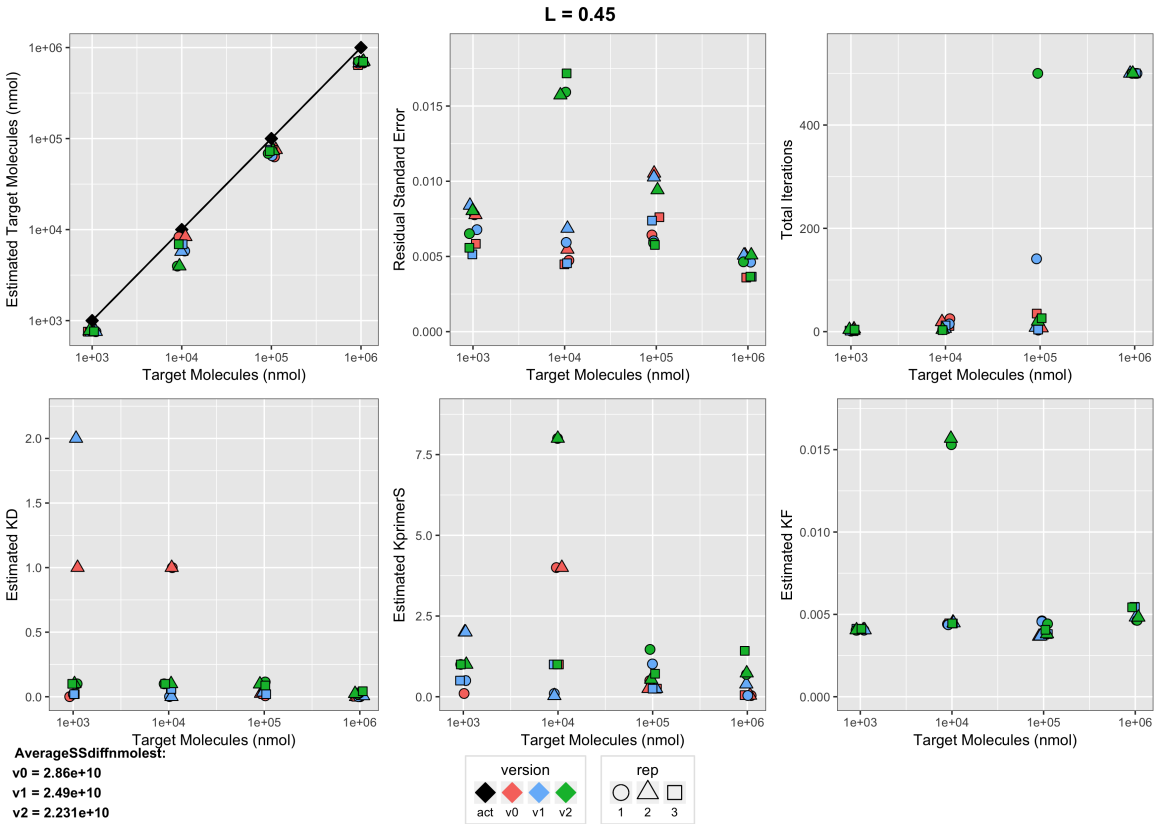


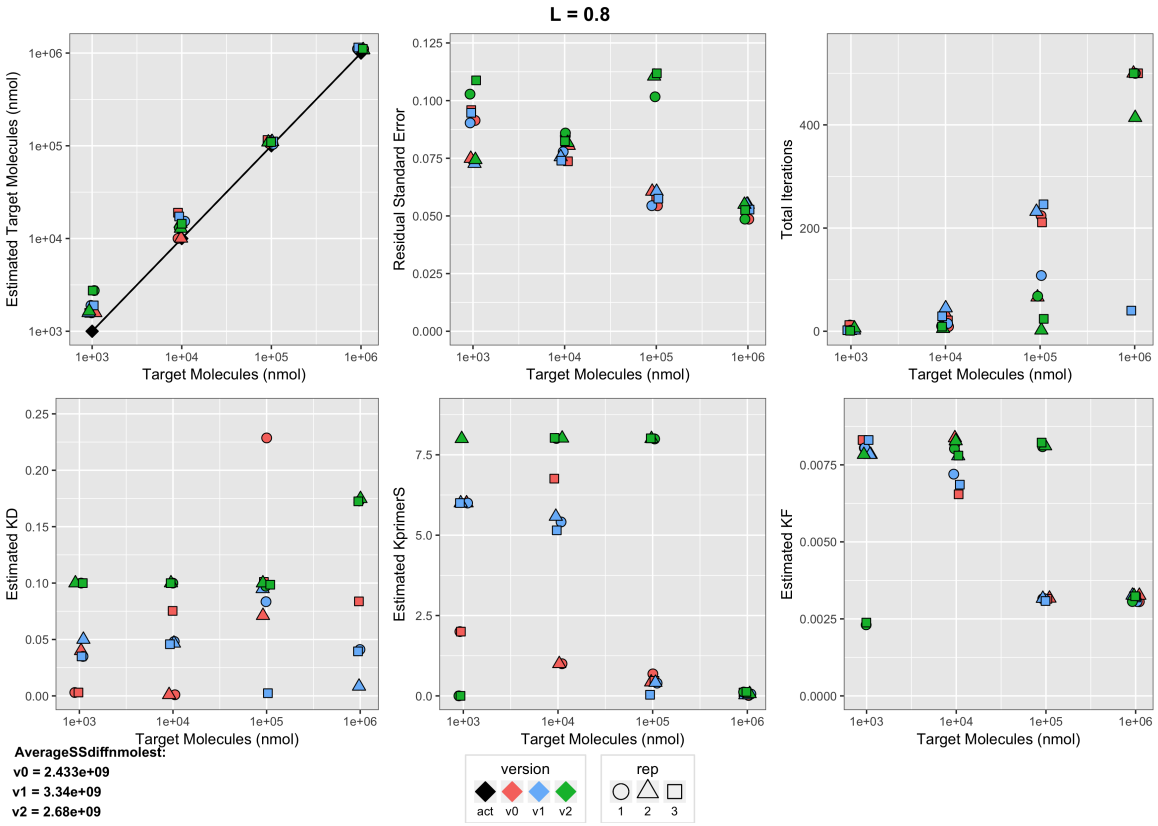
Based on these results, only the constant from the new method was used when evaluating the performance of the different versions.

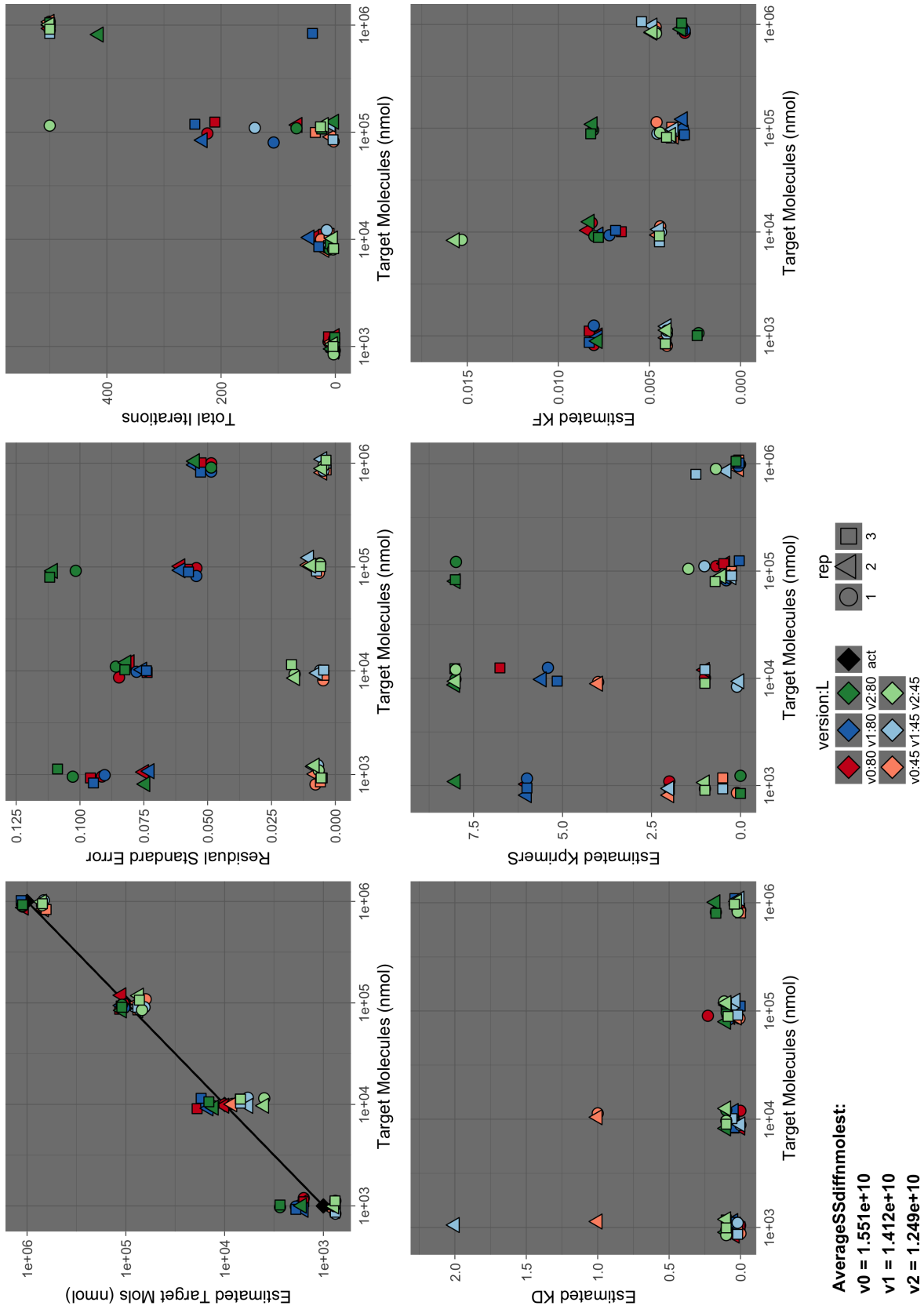
Elapsed Time

The time elapsed was not recorded for most runs. However, it was recorded for the old constant run with $L = 0.8$ for version 1 and version 2 of the code. Version 1, which estimated K_D took apprximately 4 hours and 15 minutes to complete. Version 2, where K_D was given, took approximately 30 minutes.

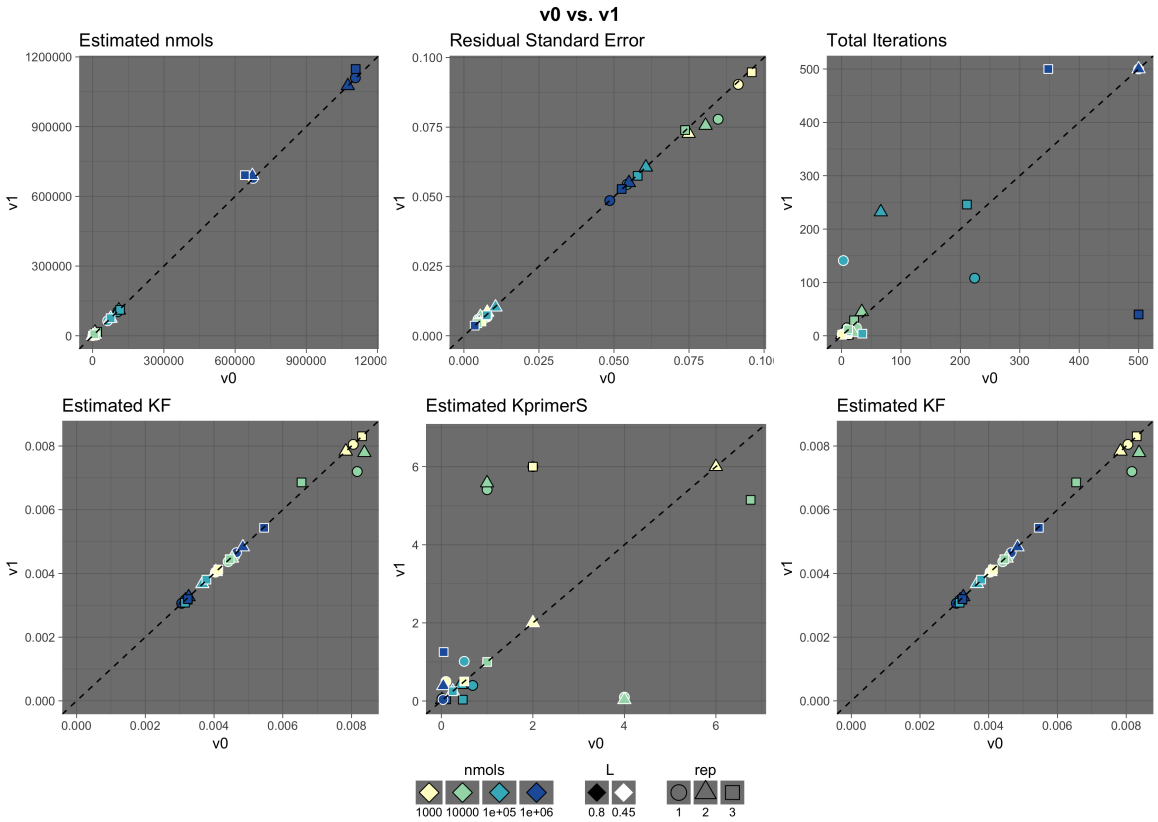
Result Plots

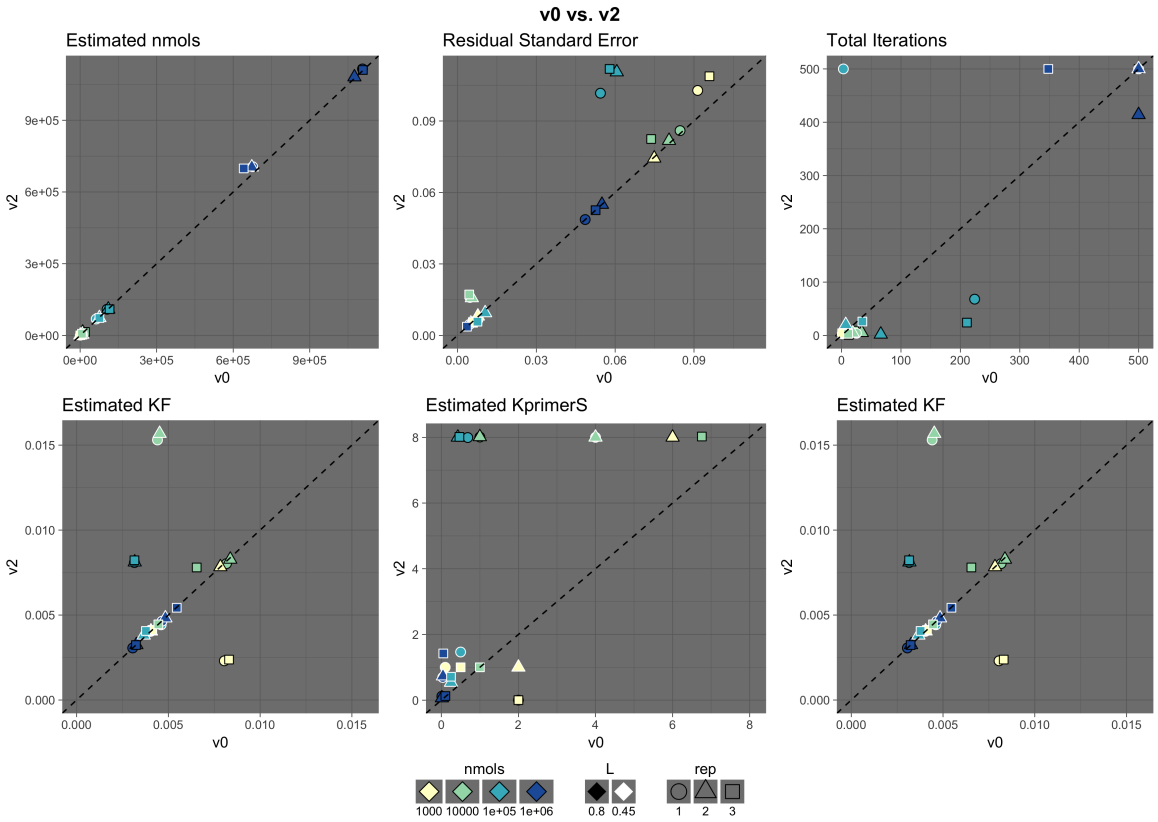


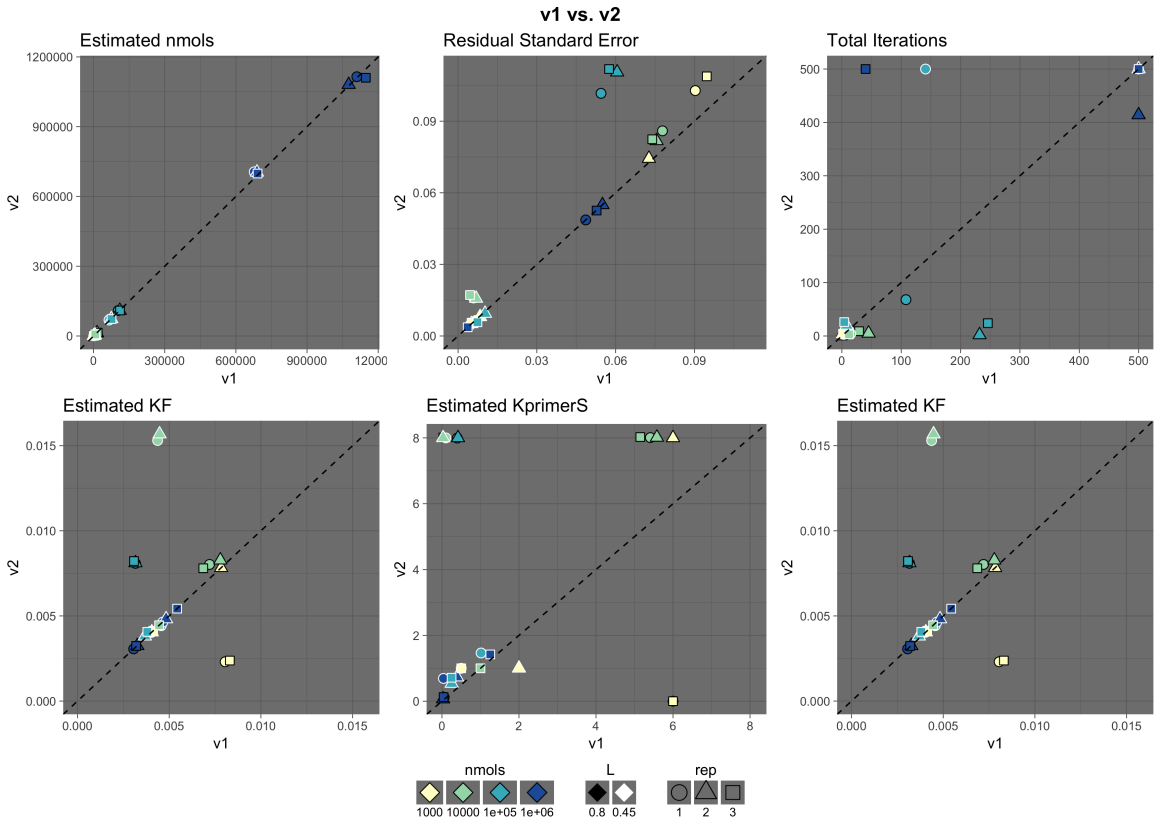



Estimated Target Mols (nmol)
Residual Standard Error
Total Iterations
Estimated KD
Estimated KPrimerS
Estimated KF
Target Molecules (nmol)
1e+03
1e+04
1e+05
1e+06
0.125
0.100
0.075
0.050
0.025
0.000
400
200
0
2.0
1.5
1.0
0.5
0.0
7.5
5.0
2.5
0.015
0.010
0.005
version:L
v0:80
v1:80
v2:80
act
v0:45
v1:45
v2:45
rep
1
2
3
AverageSSdiffnmolest:
v0 = 1.551e+10
v1 = 1.412e+10
v2 = 1.249e+10

Version Comparison



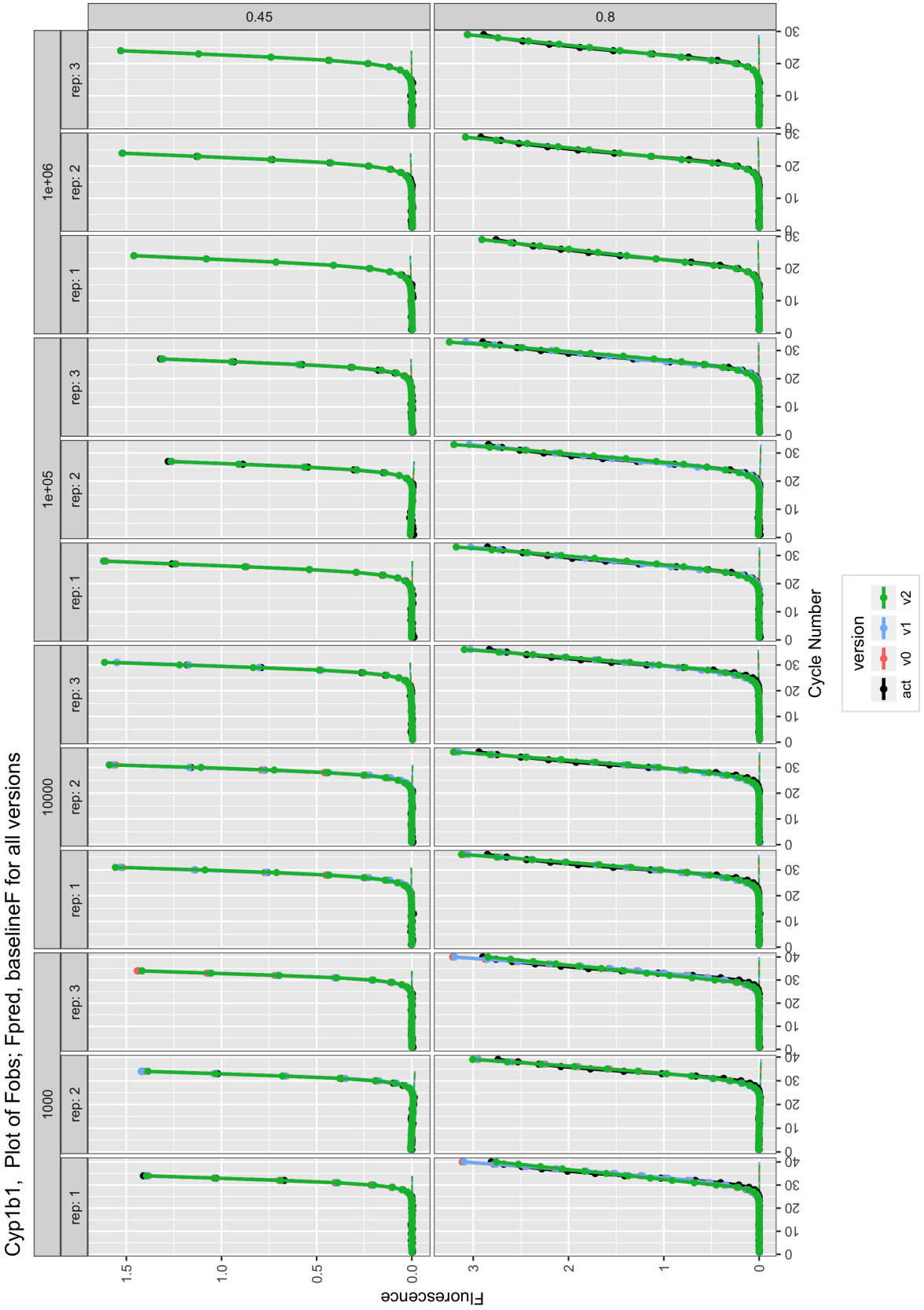


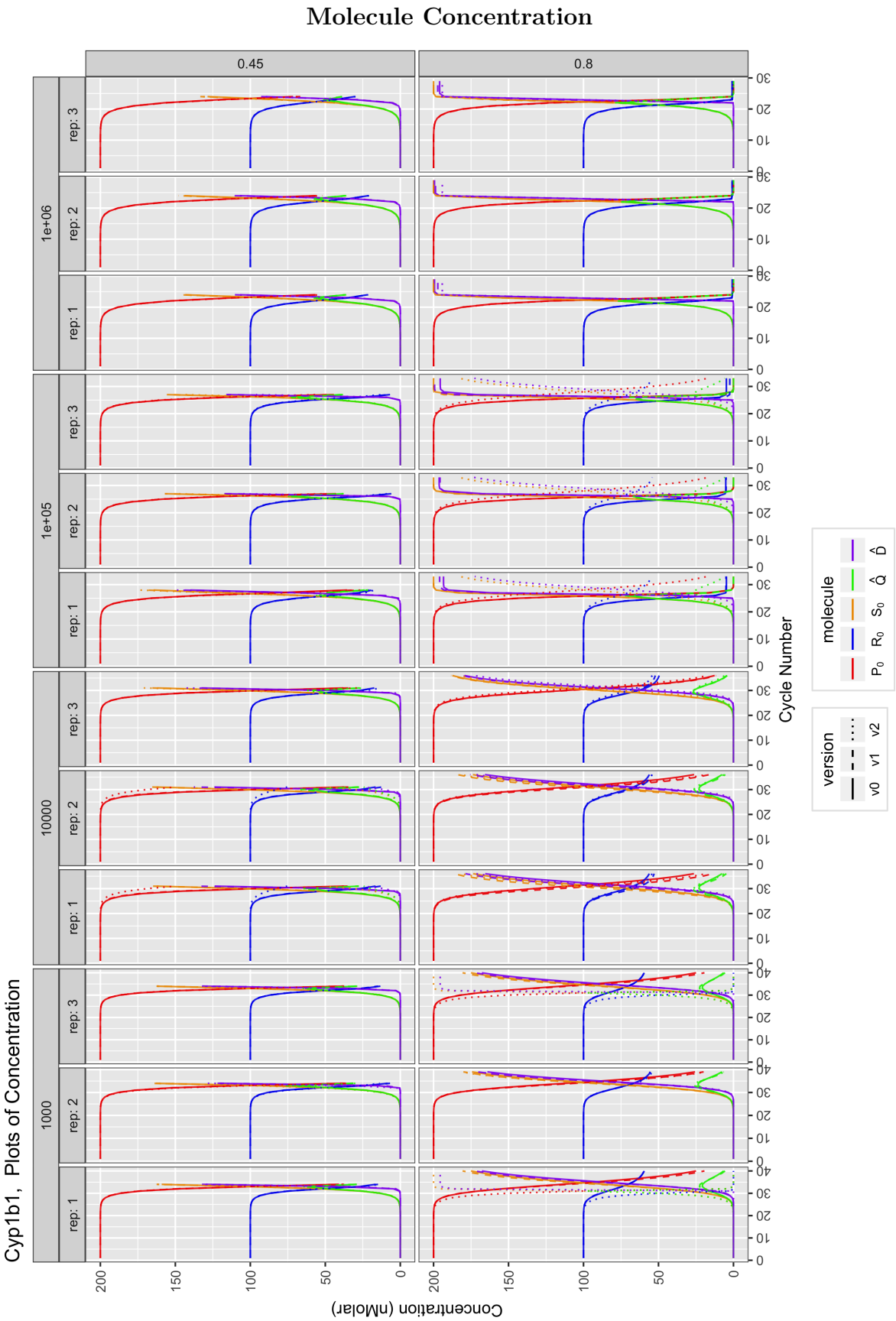


Avg Sum of Squares Difference Between NMol Est. and Actual Value

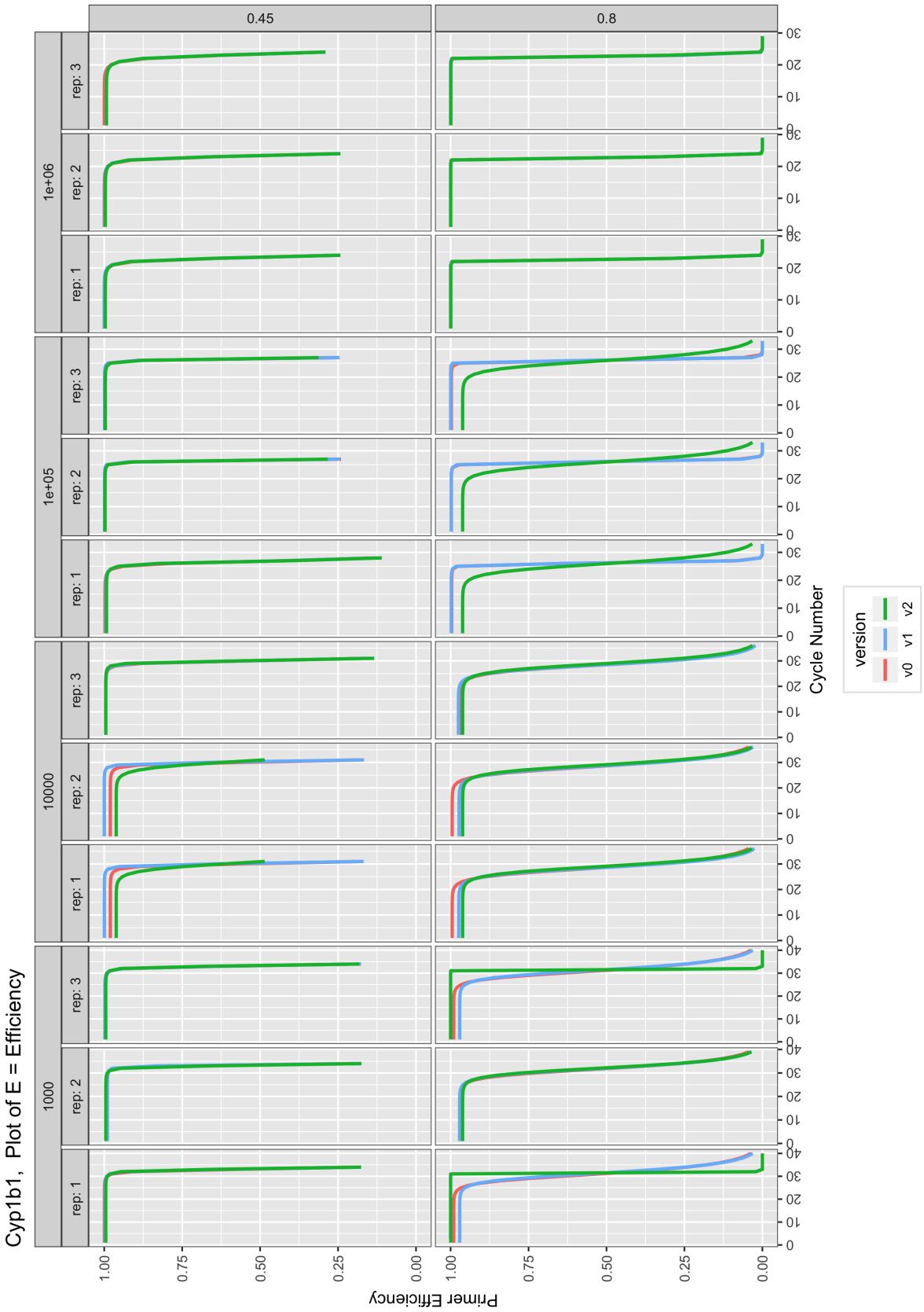
Version	L Value: **0.8**	**0.45**	**avg**
v0	2.433×10^{9}	2.860×10^{10}	1.551×10^{10}
v1	3.340×10^{9}	2.490×10^{10}	1.412×10^{10}
v2	2.680×10^{9}	2.231×10^{10}	1.249×10^{10}

Fit to qPCR Curve




Molecule Concentration
Cyp1b1, Plots of Concentration
0.45
0.8
1e+06
1e+05
10000
1000
rep: 1
rep: 2
rep: 3
Concentration (nMolar)
Cycle Number
molecule
P_0 R_0 S_0 $\hat{Q}$ $\hat{D}$
version
v0 v1 v2

Efficiency


Cyp1b1, Plot of E = Efficiency
0.45
0.8
1000
10000
1e+05
1e+06
rep: 1
rep: 2
rep: 3
Primer Efficiency
1.00
0.75
0.50
0.25
0.00
Cycle Number
version
v0
v1
v2

Conclusions

The version of the code which used the constant multiplier and given K_D value performed about the same as or better than the versions which estimated K_D and in a much shorter amount of run time.

The old constant multiplier method performs more poorly than the newer method, which suggests that the length of the primer and probe are more important than their specific sequences or the annealing temperature.

All versions of the code performed poorly when using the full, un-truncated data set. The manufacturer of the enzyme kit used in the Cyp1B1 data set suggests that the enzyme may be a limiting factor and cause for the poor fit at later cycles.(Thermo Fisher Scientific, n.d.) Including the amount of enzyme 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

Gotoh, M. (1995, Jan). A new approach to determine the effect of mismatches on kinetic parameters in DNA hybridization using an optical biosensor. *DNA Research*, *2*(6), 285-293. Retrieved from https://academic.oup.com/dnaresearch/article-lookup/doi/10.1093/dnares/2.6.285 doi: 10.1093/dnares/2.6.285

Jensen, K. K., Ørum, H., Nielsen, P. E., & Nordén, B. (1997, Apr). Kinetics for hybridization of peptide nucleic acids (PNA) with DNA and RNA studied with the biacore technique. *Biochemistry*, *36*(16), 5072-5077. Retrieved from https://doi.org/10.1021/bi9627525 doi: 10.1021/bi9627525

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

Sutter, T., Tang, Y., Hayes, C., Wo, Y., Jabs, E., Li, X., ... Greenlee, W. (1994, Jul). Human dioxin-inducible cytochrome P450 (CYP1B1) mRNA, complete cds. Retrieved from http://www.ncbi.nlm.nih.gov/nuccore/U03688.1

Thermo Fisher Scientific. (n.d.). *4304437 documents & support and product FAQs.* Retrieved from https://www.thermofisher.com/search/results?query=4304437&persona=DocSupport&type=Product+FAQs

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

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012, Jun). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics*, *13*, 134. doi: 10.1186/1471-2105-13-134