Industrially Robust Synthetic Biology Standards for the Polymerase Chain Reaction

A thesis submitted to the University College London

for the degree of

ENGINEERING DOCTORATE

by

Alexander Templar BSc MRes

2017

Advanced Centre of Biochemical Engineering

Department of Biochemical Engineering

University College London

Torrington Place

London

WC1E 7JE

Table of Contents

Abstract

Synthetic Biology is ushering in a new era where reengineered genomes can enhance the capacity of host cells to produce biologic and chemical products. Standardisation is a key component of synthetic biology as it enables effective implementation (Müller and Arndt, 2012). This project has successfully generated synthetic biology standards for the quantitative polymerase chain reaction (qPCR), a highly specific and sensitive analytical platform, in order to increase its robustness for monitoring of host cell processes in an industrial setting. This project has also increased the assay throughput to allow for at-line analysis, in accordance with initiatives such as Process Analytical Technology (PAT) (Gnoth et al., 2007). Analysis was conducted on three commonly used host cell chassis and industrial contamination was also simulated by the addition of plasmid proxies. All assays were optimised by primer design and screening to ensure accuracy. End point PCR (e –pPCR) and quantitative PCR (qPCR) was conducted in the presence and absence of cellular material disrupted by a mild sonication procedure. We found that, whilst cellular material reduces assay sensitivity for a genomic locus, the presence of contaminating species can be accurately quantified. We also employed LRE–qPCR, which uses the CAL1 standard for quantification. LRE-qPCR matched the accuracy of a conventional standard curve qPCR method and we propose it as a Synthetic Biology standard.

We next developed a modified standard curve method that streamlined methodology and bypassed errors inherent to the gold standard methodology to, for the first time, enable quantification of multiple targets from a single standard curve. The CyCal curve is a standard curve constructed from the CAL1 standard combined with the Cy0 data analysis. The approach was validated against 6 bioprocess targets and it was found that CyCal was able to replicate the accuracy of the gold standard approach. We then used CyCal to accurately determine how host cell plasmid copy number (PCN) evolves during fermentation. The combination of rapid sample preparation and a universal standard means that CyCal is capable of becoming the basis of an at-line qPCR assay when conducted on modern ultra-rapid qPCR thermocycler technology.

Thesis Declaration

I, Alex Templar, hereby declare that this thesis is my own work and effort and that it has not been submitted anywhere for any award. Where other sources of information have been used, they have been acknowledged.

Signature: …….

Date: …….

Acknowledgements

I would like to thank Darren for his patient supervision during these four years, especially for the publication of the papers from chapters 3, 4 and 5. His contributions and guidance made this project possible and his approach made it enjoyable. I'd also like to thank my family, especially my mum and dad, who have not only kept me motivated but have also prevented me from becoming bankrupt on more than one occasion.

Therefore this work is dedicated to them, my Mum and my Dad, as I couldn't have achieved this without their support. I love you both.

Nomenclature

Abbreviations

Bioreactor	BR
BSM	Basal Salt Media
BMGY	Buffered glycerol complex medium
CHO	Chinese hamster ovary
Cq	Cycle of quantification
DOE	Design of experiments
dsDNA	Double stranded DNA
(DMEM)	Dulbecco's Modified Eagles Media
e-pPCR	End-point PCR
Ft	Fluorescence threshold
Fab	Fragment antigen binding fragment
gDNA	Genomic DNA
GS	Glutamine synthetase
HCD	High cell density
IgG4	immunoglobulin G4

LOD	Limit of Detection
LCD	Low cell density
LRE-qPCR	Linear Regression of Efficiency qPCR
LB	Luria Broth
mAb	Monoclonal Antibody
OCF	Optical Conversion Factor
OD	Optical density
PTM1	Pichia Trace Metal 1
PCN	Plasmid copy number
PCR	Polymerase Chain Reaction
PAT	Process analytical technology
qPCR	Quantitative PCR
YPD	Yeast extract peptone dextrose
SC PCR	Standard Curve PCR
SF	Shake Flask
SEAP	Secreted alkaline phosphatase

ssDNA	Single stranded DNA
TKT	Transketolase
wcw	Wet cell weight
WCB	Working cell bank

Chemicals and Elements

$(NH_4)_2SO_4$	Ammonium Sulphate
$C_6H_8O_7$	Citric Acid
$CaCl_2.6H_2O$	Calcium chloride hexahydrate
$CaSO_4$	Calcium Sulphate
$CoCl_2$	Cobalt (II) Chloride
$CoSO_4.7H_2O$	Cobalt(II) Sulfate Heptahydrate
$CuSO_4$	Copper(II) Suphate
$CuSO_4.5H_2O$	Copper(II) Suphate Penthydrate
$FeCl_3.6H_2O$	Iron Chloride Hexahydrate
$FeSO_4$	Iron(II) Sulphate
H2O	Water

H2PO4	Dihydrogen phosphate
H2SO4	Sulphuric acid
H3BO3	Boric Acid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K2SO4	Potassium sulphate
KCl	Potassium chloride
KOH	Potassium hydroxide
MgSO4	Magnesium sulphate
MgSO4.7H2O	Magnesium sulphate heptahydrate
MnSO4	Manganese sulphate
MnSO4.4H2O	Manganese sulphate tetrahydrate
Na2MoO4.2H2O	Sodium molybdate dihydrate
NaCl	Sodium chloride
NaH2PO4.H2O	Sodium dihydrogen phosphate
Nal	Sodium iodide
ZnCl2	Zinc chloride
$ZnSO_4.7H_2O$	Zinc sulphate heptahydrate

List of Figures

List of Tables

Chapter 1

1 General Introduction

1.1 Introduction to synthetic biology and standardisation

Synthetic biology is an engineering-driven approach to utilising recombinant DNA technology that has proved to be immensely successful in developing an understanding of how genetic systems and metabolic pathways function (Arpino et al., 2013). Synthetic biology has allowed entirely synthetic (Gibson et al., 2010) or refactored (Lajoie et al., 2013) genomes to be used to design new function into organisms such as bacteria and has been crucial in the development of increasingly sophisticated methods of manipulating of genetic systems (Agapakis and Silver, 2009).

Standardisation, or the universal adoption of conventions, materials or methodologies, is key to the synthetic biology ethos as it facilitates communication, compatibility, quality control and reproducibility across this highly multidisciplinary field (Müller and Arndt, 2012). By adopting standards experiments can become comparable across scientific disciplines as data can be shared and results reported with greater ease. Standardisation in Synthetic Biology is largely defined by a "toolbox" of well characterised genetic components (Kelly et al., 2009), organisms and protocols (Bustin et al., 2009).

1.2 The quantitative polymerase chain reaction

The Polymerase Chain Reaction (PCR) is an *in vitro* method of amplifying a specific target region of DNA. The process involves DNA polymerase-mediated reactions in which DNA amplification is driven by a series of temperature cycles, with a full PCR run consisting of 30 cycles or more. The DNA products of PCR are termed amplicons and a perfectly efficient reaction will double the quantity of amplicons with each cycle (Erlich et al., 1991).

Amplification by PCR can be observed in real-time by introducing fluorescent probes into the reaction media. These probes react with double stranded DNA (dsDNA) to

produce a signal proportional in magnitude to amplicon quantity, making it possible to track the production of amplicons in real-time (Wilhelm and Pingoud, 2003). This real-time gathering of data confers a major advantage as the reaction container can remain closed once an experiment has begun, reducing the risk of sample contamination and improving experimental reliability. The need for complex downstream processing is also eliminated and a high degree of automation can be applied, which in turn reduces experimental time and cost and allows for higher-throughput assays (Kubista et al., 2006), (Wilhelm and Pingoud, 2003).

qPCR was pioneered by Higuchi et al. (1992), the first group to record the increase in fluorescent signal as reporter molecules interact with exponentially accumulating amplicons (Higuchi et al., 1992). It was quickly realised that kinetic PCR data, in the form of an amplification profile, can be subjected to powerful quantitative analyses that are capable of providing information on original template mass in samples where it was previously unknown (Wilhelm and Pingoud, 2003). qPCR has since achieved "gold standard" status due to the myriad of other convenient experimental factors, such as the improved experimental reliability as a result of data gathering from a closed reaction container, the reduced need for complex downstream processing and the amenability of qPCR to automated and high-throughput systems (Kubista et al., 2006).

1.2.1 PCR nomenclature

The rapid and widespread uptake of qPCR has resulted in varying nomenclature within the literature. There are a number of terms that, whilst broadly describing the same process, have subtly different meanings. Real-time PCR refers to a PCR where the data is gathered whilst the amplification of DNA is occurring. Kinetic PCR also refers to this process and is fully interchangeable with real-time PCR. Quantitative PCR is used to describe a PCR reaction where the amount of starting DNA can be quantified. Whilst the term "quantitative PCR" has been used in the literature before the development of real-time PCR (Ferre, 1992) the two terms are now interchangeable, due to the widespread use of real-time PCR as a quantitative technique.

Real-time PCR is sometimes shortened to RT-PCR in the literature; however RT-PCR is also used to describe reverse-transcription PCR, a form of PCR where reverse transcriptase enzymes are used to convert RNA before the PCR. Due to the confusion between reverse-transcription PCR and real-time PCR, and the common use of quantitative PCR to describe real-time PCR, qPCR will be used as the term to describe real-time PCR, kinetic PCR and quantitative PCR.

1.2.2 PCR data capture and analysis

As PCR protocols and methodologies have evolved the methods of acquiring data from the process of nucleic acid amplification have diversified. There are currently two main methods of studying the products of PCR, end-point analysis and through a qPCR amplification curve.

1.2.3 End-point PCR

In end-point PCR (e-pPCR) the amplified DNA is analysed at the end point of the reaction using gel electrophoresis. Although the quantification of target DNA is possible with densitometry, e-pPCR is often used as a binary analysis method, giving a "yes" or "no" result as to the presence of a target gene and allowing visualisation of amplicon size (Aaij and Borst, 1972).

Whilst e-pPCR has been superseded by other PCR methods, such as qPCR, it is still used to detect clinical or industrial infection by adventitious microorganisms. These can include mycoplasmas, phage and other contaminating bacteria (Kong et al., 2001; Arabestani et al., 2011). Through elegant assay design, primers can be designed that are capable of distinguishing between multiple contaminant species in a single sample (Timenetsky et al., 2006).

1.2.4 qPCR amplification curve

Kinetic qPCR data is plotted as an amplification profile, shown in Figure 1-1. The amplification profile has distinct stages; the linear-ground, exponential and plateau

phases, plus inter-phases between them. These reflect the processes behind amplicon production and accumulation, with the linear-ground phase reflecting the cycles during which fluorescence from product accumulation has not exceeded background fluorescence (F_b) (Ruijter et al., 2009), the exponential phase reflecting the cycles where the increase in amplicon mass reaches an exponential level according to the growth equations (Rutledge and Stewart, 2008), and the plateau phase reflecting the stalling of the reaction, largely due to inhibition through excess product accumulation (VanGuilder et al., 2008).

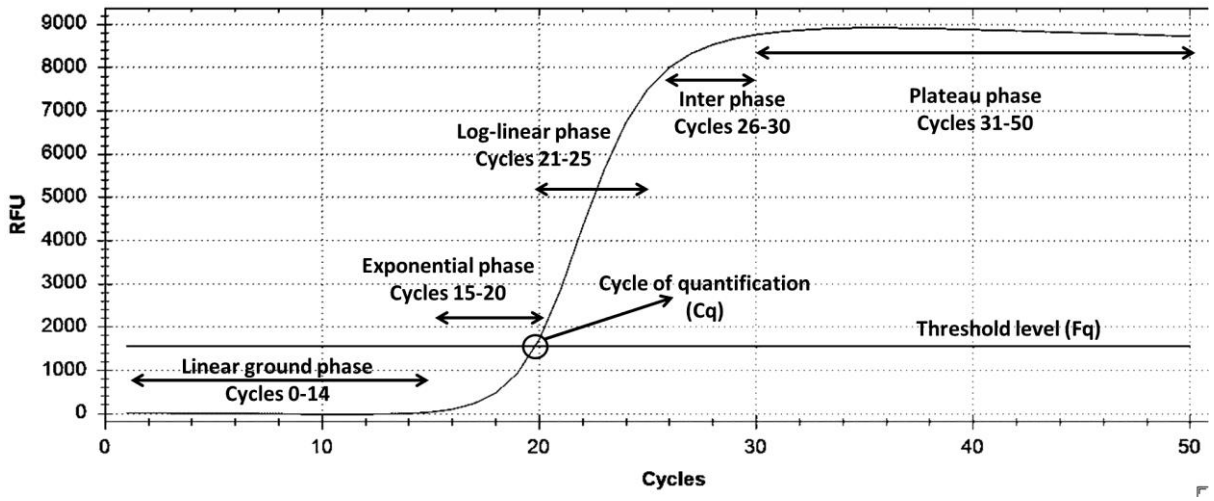

Figure 1-1 qPCR amplification curve annotated with kinetic phases

1.2.5 Efficiency of qPCR amplification

The accumulation of amplicons during PCR can be described by an application of the exponential growth equation, which establishes a relationship between the efficiency (E) of amplification and quantity of nucleic acid template (N):

$$N_n \cong N_0 E^n$$

Where N_n is the quantity of target DNA at cycle n and N_0 is the quantity of target DNA initially input to the reaction. E can range between values of 1, which describes 0% efficiency where no amplicons are produced, or 2, which describes 100% efficiency where amplicons double during each cycle (Tichopad et al., 2003).

As it has been shown that fluorescent intensity is proportional the quantity of template, values for fluorescence signal can be directly substituted:

$$F_n = F_0 E^n$$

Where F is the fluorescence signal from the nucleic acid target, Fn is fluorescence at cycle n, and F_0 is the initial fluorescence. As fluorescence intensity is directly proportional to initial template quantity, F_0 also represents the absolute quantity of template (Rutledge and Côté, 2003).

These calculations are the basis for nucleic acid quantification from the fluorescent signal generated during qPCR and demonstrate that accurate calculation of E is critical to the quantitative precision of qPCR.

There are numerous compounds that can impact efficiency of qPCR amplification, inhibiting amplicon formation and causing quantitative inaccuracy. To maintain the accuracy of assays, it is common to subject a test sample to round or rounds of sample preparation to remove these inhibitors and allow for testing on a sample of the highest possible purity.

1.2.6 Quantification cycles (Cq) in qPCR

As efficiency has a direct impact on quantification, analysis is carried out in the exponential phase of amplification where efficiency is considered to be constant (Ruijter et al., 2009). This is achieved through generation of the cycle pf quantification (*Cq*), or the fractional cycle at which fluorescence intensity exceeds the fluorescence threshold (*Fq*), a line drawn within the exponential phase, above the background fluorescence (*Fb*), which runs parallel to the x-axis of the amplification curve (Ruijter et al., 2009). *Fb* is the level of fluorescence generated from background sources such as the experimental sample, reagents and qPCR consumables being used. The *Fq* is commonly generated by the qPCR instrumentation, but can be implemented by the user. If implemented by the user then the *Fq* should be set at the same point for all experimental repeats to ensure comparability (Ruijter et al., 2013).

1.2.7 Relative quantification from qPCR

Relative quantification compares the change in levels of a target nucleic acid sequence to an experimental control group, giving the fold difference in gene expression. Absolute quantification gives a measure of the absolute quantity of nucleic acid and typically involves the construction of a standard curve (SC), where samples containing a known mass of template are amplified in parallel, to gain information on a sample containing the same target sequence but unknown mass (Rutledge and Côté, 2003).

1.2.8 Absolute quantification qPCR

Whilst direct observation of F_0 would conveniently achieve absolute template quantification, it is unfortunately rendered impossible by interference from F_b. Therefore F_0 is approximated using information from the exponential phase of the amplification profile, where amplicon fluorescence suitably exceeds F_b.

The analysis of the exponential phase of amplification is achieved through generation of the cycle of quantification (*Cq*). This is the fractional cycle at which fluorescence

intensity exceeds the fluorescence threshold (F_q), a line drawn within the exponential phase that runs parallel to the x-axis of the amplification curve. Some qPCR instrumentation, such as the Roche LightCycler, can also calculate Cq using the second derivative maximum (SDM), identified as the point on the amplification curve at which the increase of fluorescence intensity reaches its maximum (Pfaffl, 2001).

The SC method takes 10 fold serial dilutions of a sample of known DNA mass and amplifies them in parallel to generate a series of Cq values. These are then plotted against their log dilution to give a linear regression line, of which the y-intercept (b) and slope (m) are used to convert the Cq of an unknown sample into quantified template mass, using the following equation:

$$N_n = 10^{\left(\frac{C_q - b}{m}\right)}$$

Most commercially available packages for qPCR data analysis apply baseline subtraction, or a process of accounting for F_b to improve quantitative precision.

The increased labour needed in setting up a qPCR assay capable of absolute quantification is a contributing factor, as SC protocols increase experimental complexity and error from the extra set of liquid handling steps. The most important issue however is that assumptions are made during data analysis that reduces quantitative precision.

1.2.9 Limitations of absolute quantification

A fundamental assumption made by the SC method is that the purified standards and sample of unknown DNA mass have an equal value for amplification efficiency, E, which has been experimentally proven to be highly improbable (Rutledge and Stewart, 2008). As previously discussed, accurate and consistent determination of E is critical to the quantitative precision of qPCR. Differences in E of only 0.05 (corresponding to 5%) can result in a sample having double the mass of amplicons

after 26 cycles of amplification (Freeman et al., 1999). The implications of this are that even small differences in E between standard and unknown sample efficiencies often result in loss of quantitative accuracy and, in extreme cases, erroneous data (Meijerink et al., 2001). Some studies even suggest that E is a dynamic value, in that its value varies within a cycle (Goll et al., 2006).

1.3 Diversity of qPCR data analysis

A number of alternate protocols have been developed that aim to achieve accurate absolute quantification without the practical flaws and inherent assumptions of the standard curve method. Amongst these are attempts to give a more reliable and dynamic estimation of E, to provide a more accurate determination of Cq and to model the mechanics of amplicon accumulation to extrapolate initial template fluorescence and mass.

The ultimate goal of these refinements in qPCR data analysis is a method of accurately determining F_0, and therefore initial template mass, from a single reaction with minimal assumptions. This would allow for accurate and rapid absolute quantification that is highly amenable to automation. The development of a superior method of absolute quantification could also drive a synthetic biology level of standardisation, which would see diverse methods and arbitrary experimental decisions coalesce into a routine laboratory assay that produces consistently reliable data.

1.3.1 Sigmoidal curve fitting and Linear Regression of Efficiency (LRE)-qPCR

Sigmoidal curve fitting was the first attempt to estimate F_0 from observations of raw fluorescence data (Liu and Saint, 2002) and since its appearance in the literature the method has been extensively built upon, resulting in LRE-qPCR (Rutledge, 2004). By modelling the qPCR amplification curve with a Boltzmann sigmoid function, this method attempts to calculate efficiency and fluorescence values at specific cycles (Rutledge, 2004). The sigmoid function that was first applied to qPCR data is:

$$F_n = \frac{F_{max}}{1 + e^{-\left(n - n_{1/2}/k\right)}} + F_b$$

Where n is cycle number, F_n is fluorescence at cycle n, F_{max} is the maximum value of fluorescence at the end of PCR, $n_{½}$ is the fractional cycle at which fluorescence reaches half of F_{max}, K is the slope of the curve and F_b is background fluorescence. Following calculation of F_c, the exponential growth equation can be further applied to values generated by the model to estimate F_0 and thus the initial template mass.

Whilst this approach is successful at modelling the early log-linear stages of amplification, the sigmoid function deviates from fluorescence values found in the plateau phase and becomes inaccurate, thus making it is necessary to find a "cut-off" point after which fluorescence values are ignored. This was initially chosen by the experimenter, however LRE-qPCR developed by Rutledge implemented an automated method for choosing a suitable cut-off point. This automated process uses an algorithm to determine the cycle where the increase in fluorescence reaches its lowest value and then removes all subsequent plateau phase cycles, producing an analytical window that only contains values in the linear phase of amplification (Rutledge and Stewart, 2010).

Subsequent developments in LRE-qPCR further improved quantitative accuracy by introducing an improved method of converting the raw fluorescence value, F_c, into quantities of nucleic acid. This was achieved by establishing a standardisable optical scale, or optical conversion factor (OCF), to convert target fluorescence into target copy number, negating the need for target specific calibration. A high performance calibration reaction using commercially available lambda DNA is suggested and details of the calibration primers and reaction conditions have been published and are openly available(Rutledge and Stewart, 2010).

A javascript program, which conveniently automates the application of LRE-qPCR to raw fluorescence values generated from qPCR, has been developed and is maintained, with an OCF calculation feature included (Rutledge, 2011).

Studies have demonstrated that LRE-qPCR can give quantitative accuracy of ±15-25% from "gold standard" standard curve methods and it is also been demonstrated that single-molecule template reactions can be analysed with this method (Rutledge and Stewart, 2010). Additionally, the calibration done with lambda DNA allows for a degree of standardisation. Some reports have however concluded that this sigmoidal curve fitting method assumes symmetry between the inflection points between the ground-leading-to-exponential and exponential-leading-to-plateau phases of the amplification curve, introducing errors in quantitation and meaning that the standard curve (SC) method remains gold standard (Ruijter et al., 2013).

1.3.2 Optical Calibration Factor (OCF) and LRE-qPCR

LRE-qPCR allows determination of F_0 through analysis of an amplification curve, which must then be converted into the number of target molecules by correlating the fluorescence units to the mass of the DNA. In order for this to be practical, a correlation must be established between the DNA mass and fluorescence. This can be achieved by amplifying a known quantity of standard with as close to 100% efficiency as possible. This approach is similar to conventional qPCR approaches, with the exception that it doesn't necessitate specificity to the target molecule. Rutledge et al (2008) have developed a "calibration reaction" for this, which uses commercially available lambda bacteriophage genome as target and a high-performance 'Cal1' primer pair.

1.3.3 Cy0

Cy0 is a modified standard curve method that produces a Cy0 value that, unlike the Cq value, does not require the assumption of uniform e between the standards and the unknown sample and does not require an arbitrary F_q.

The Cy0 is both a quantitative entity and a reaction kinetic, and was realised through adaptations of the logistic models described by Rutledge & Stewart (Rutledge, 2004) and Chervoneva et al (2007). The Cy0 value is generated by using non-linear regression of the experimental data to produce a Richards curve. The tangent of the inflection point

of this curve is intersected with the abscissa axis, the point of intersection being the Cy0 value. The 5 parameter Richards function used is:

$$F_n = \frac{F_{max}}{\left[1 + E\left(-\frac{1}{b}(n-x)\right)\right]^d} + F_b$$

Where F is the reaction fluorescence at cycle n, which is the fractional cycle of the turning point of the curve, b is the slope parameter, x is the transition parameter and d is the Richards coefficient.

If d=1 then the Richards equation becomes the logistic equation and the parameters obtained by these equations can be used to generate the Cy0 value (Guescini et al., 2008):

$$C_y0 = x - b(1 + \frac{1}{d} - \ln d)$$

Since publication of the original Cy0 method, the authors have integrated shape based kinetic outlier detection (SOD) into analysis that compensates for variations in reaction efficiency between samples and improve accuracy (Guescini et al., 2013).

The authors of Cy0 attempted to prove their model by simulating varying levels of PCR inhibition through the addition of Immunoglobulin G (IgG), a known PCR inhibitor. Two criteria were used to judge performance, accuracy of quantification and variability between results. They found that, whilst Cy0 and SC methods have comparable levels of variability and accuracy when amplification conditions were optimal, the accuracy of SC methods became significantly impaired upon the introduction of inhibition. Cy0 had levels of variability comparable to optimal conditions and had far increased levels of accuracy when the reaction was inhibited. This makes Cy0 data analysis a good choice for the absolute quantitation of samples that might contain a source of inhibition, such as those from a biological source (Ruijter et al., 2013).

The Cy0 method can be applied to data by uploading fluorescence readings to a website maintained by the authors (http://www.cy0method.org/).

1.3.4 Methods for comparing competing analysis methods

The direct quantification of target nucleic acid quantity is not currently possible, rather it is extrapolated from analysis of PCR amplification in the presence of a fluorescent reporter. Therefore, when assessing the suitability of a new data analysis method, the method will need to be evaluated through comparison to existing methods rather than the actual value of nucleic acid quantity.

Two methods of indirectly measuring a variable can be compared through an XY plot to find correlation, or the degree as to which they are related. Here the mean of one analytical method is plotted on the y axis and the mean of the second is plotted on the x axis. Visual analysis of the plots formed can give an indicator as to the nature and strength of agreement. If relationship between the variables is linear, with no noticeable curvature, linear regression analysis can be used to determine the slope and intercept of a line that best predicts Y from X, and the correlation coefficient (R^2). If the data being compared is in perfect agreement, the slope will equal 1.0 and the y intercept will equal 0. Lines with a slope not statistically different from 1.0 but with y-intercept different from 0 indicates a systematic bias. Alternatively, a slope that is statistically different from 1.0 indicates a proportional bias. R^2 is the measure of goodness-of-fit of linear regression and ranging between 1.0 (a perfect fit) and 0.0 (no linear relationship). Therefore the closer the R^2 value is to +1.0, the stronger the relationship is.

When assessing the quantitative ability of two sets of measurements, using the correlation between the methods can be misleading as it studies the relationship between the two and not the differences. Bland and Altman (1986) devised a method of comparing clinical diagnostic methods, which also indirectly measure a value, in order to assess agreement. A Bland-Altman plot determines limits of agreement, calculated using mean and standard deviations of the difference between the measurements of each method. To do this, an XY plot is constructed where the X axis shows the mean of the

two measurements and the Y axis shows difference between measurements. If the analytical methods are carried out over a large range of concentrations then the data should be log transformed. This results in difference plot, where the relationship between the averages and differences of the measurements and will reveal any bias. Should 95% of the data points lie within two standard deviations of the mean difference, which is the recommended limits of agreement (Bland and Altmann, 1986) and can be marked on the graph for avoidance of doubt, then there statistically no evidence of bias.

1.3.5 qPCR instrumentation

The development of thermal cyclers with integrated fluorimeters has been crucial to the adoption of qPCR technology and, since the first generation of qPCR thermocyclers, the instruments have become increasingly sophisticated. Early qPCR thermal cyclers that utilised a heat block to alternate between temperatures lacked homogeneity of heat within the reaction vessel, which often resulted in inefficient reactions (Mackay, 2004). Reaction efficiency is a critical component of qPCR quantification and changes in its value can dramatically affect the accuracy of the technique. In order to drive efficient reactions that result in accurate qPCR, alternative methods of heating have been developed which are able to produce a homogenous temperature gradient. These novel heating methods generally use smaller sample sizes, which in turn decreases cycle time, resulting in the development of rapid cycle qPCR (Kaltenboeck and Wang, 2005).

1.3.6 Rapid-cycle qPCR platforms

The first “rapid-cycle” qPCR instrument, marketed by Roche as the LightCycler, passed heated air over glass capillaries containing samples to drive temperature change. Thermal transfer via air is rapid and uniform, and since the commercial success of the LightCycler, competing instruments that use similar technology have entered the market (Rasmussen, 2001). Generations of qPCR thermocyclers that have followed since have improved upon efficiency and accuracy of replication. The overall cost of the instruments has also decreased, making them accessible to a larger audience and allowing them to be introduced to a wider range of applications.

The Cockerill lab published a comparison of assay turnaround time between conventional PCR and rapid-cycle PCR. They found that whilst conventional PCR had an assay turnaround time of 2 to 3 days, rapid-cycle PCR had a vastly improved turnaround time of approximately 30 minutes. It was also noted that sample contamination was markedly decreased, experiments were less laborious to perform and the range of nucleic acid quantification was approximately double that of conventional PCR (Cockerill, 2003). These factors make rapid-cycle qPCR an ideal platform for reliable, accurate assays with fast turnover time.

Table 1: Summary of commercially available qPCR thermal cyclers

Instrument	**Manufacturer**	**Probes used**	**Number of Samples**	**Sample volume (µl)**	**Capable of Rapid-cycling**
LightCycler 1.0	Roche Applied Science	Hybridisation probes, molecular beacons, Taqman probes	32	10-20	Yes
LightCycler 2.0	Roche Applied Science	Hybridisation probes, molecular beacons, Taqman probes	32	10-100	Yes
SmartCycler II	Cepheid	Molecular beacons, Taqman probes	16	25-100	Yes
Rotor-Gene 3000	Corbett Research	Molecular beacons, Taqman probes	72	10-150	Midrange
Prism 7000	Applied Biosystems	Molecular beacons, Taqman probes	96	n/a	No
Prism 7300	Applied Biosystems	Molecular beacons, Taqman probes	96	n/a	No

Prism 7500	Applied Biosystems	Molecular beacons, Taqman probes	96	n/a	No
Prism 7900ht	Applied Biosystems	Molecular beacons, Taqman probes	384	n/a	No
MyiQ	BioRad	Molecular beacons, Taqman probes	96	15-100	No
iCycler iQ	BioRad	Molecular beacons, Taqman probes, hybridisation probes	96	50	No
Mx4000	Stratagene	Molecular beacons, Taqman probes	96	10-50	No
Mx3000p	Stratagene	Molecular beacons, Taqman probes	96	25	No
Chromo4	MJ Research	Molecular beacons, Taqman probes	96	10-100	No
Opticon	MJ Research	Molecular beacons, Taqman probes	96	10-100	No
SynChron	Genetic Discovery Technology	Molecular beacons, Taqman probes, hybridisation probes	6	10-50	Yes

1.4 Introduction to Bioprocess Engineering

Bioprocess engineering (bioprocessing) is a multidisciplinary specialisation of chemical engineering that researches and develops manufacturing processes from biological

sources. These are often host cell systems that have been genetically modified using recombinant DNA technology and are grown to a high cell density (HCD) in defined fermentation conditions. A notable development in this field was the introduction of human insulin derived from transgene expression in *E. coli* in 1982 (Johnson, 1982), which was the first biopharmaceutical product bought to market. The biopharmaceutical market is now worth in excess of $50billion (Pavlov et al., 2004) and protein-based biopharmaceutical products produced from recombinant DNA technology are projected to represent 19-20% of the total market value by 2017 (IMS Health, 2016). This rapid growth in the biopharmaceutical market, as well as the increased production of biofuels (Clomburg and Gonzalez, 2010), biocatalysts (Ruth and Glieder, 2010) and other fine chemicals (Zhang and Wang, 1994) from recombinant DNA technology, means that bioprocessing as a transformative research field is likely to endure.

1.5 Bioprocess platforms

Since *E. coli,* the first bacterial host cell chassis, bought recombinant insulin to market, many other cellular bioprocess platforms have been explored. A significant portion of the biopharmaceutical market utilises eukaryotic hosts, primarily mammalian cells such as the Chinese hamster ovary (CHO), but also yeasts such as *Pichia pastoris*. Bioprocessing is continuing to produce an ever increasing array of compounds from these organisms, as well as optimising them to improve product titre from batch production and increase product quality (Riesenberg and Guthke, 1999).

1.5.1 Pichia pastoris

The methylotrophic yeast *P. pastoris* is a eukaryotic industrial chassis with a proven capacity for expressing hundreds of different recombinant proteins that has emerged as an effective microbial cell factory for production of biotherapeutics (Macauley-Patrick et al., 2005) and bespoke fine chemicals (Zhang and Wang, 1994). *P. pastoris* possesses some advantages over mammalian expression systems, such as a greater amenability to genetic modification (Lin Cereghino et al., 2001) and ability to rapidly reach and maintain high cell densities (HCD), whilst retaining the capacity for human-compatible post translational modification (Sreekrishna et al., 1988).

Challenges for *P. pastoris* bioprocessing include mitigation of the impacts of HCD cultivation on cell physiology (Heyland et al., 2010) and the downstream processing challenges posed by process stream with a high content of solids, such as dewatering (Lopes et al., 2012) and clarification (Tolner et al., 2006). Developing optimal *P. pastoris* chassis and process design therefore remains an active area of investigation for bioindustry.

1.5.2 Chinese Hamster Ovary (CHO)

The first therapeutic protein produced within a mammalian cell host was a human plasminogen activator in 1986 (Vehar et al., 1986) and, since this, mammalian cell production platforms have now become the primary means of commercially producing recombinant proteins (Walsh, 2014). Mammalian cells have the capacity to provide human-like post glycosylation patterns, a feature essential for effective production of monoclonal antibodies (mAbs) that represent more than half the products in the biopharmaceutical industry. Due to features such as high genetic plasticity and adaptability to different culture conditions, the Chinese Hamster Ovary (CHO) has become one of the most widely researched and used mammalian host cell chassis for the production of mAbs (Jayapal et al., 2007). Furthermore, the use of CHO cells over two decades has demonstrated that they are safe and reliable hosts, increasing the probability of regulatory approval of associated products.

In their 2 decades of CHO cell line use, an increase in productivity of more than 100-fold have been observed (Hayduk and Lee, 2005). This has been due to optimisation strategies and the development of serum free media (Hacker et al., 2009). The demands of an escalating and competitive market mean that increase in product titre is still an active area of research and strategies of increasing cell density and cell productivity are under development (Kantardjieff et al., 2009).

1.5.3 Escherichia coli

The gram negative bacterium *Escherichia coli* has seen the longest use as a chassis for the commercial production of recombinant proteins. It remains a promising host for the

production of biologics due to its rapid rate of growth and cultivation, easy and well-described genetic manipulation and relatively inexpensive cultivation tools and media. A large body of research has been carried out detailing the production of recombinant protein in *E. coli* and groups have achieved the production of up to 285kDa proteins with gram-per-litre yields from HDC culture (Xia et al., 2010, Huang et al., 2012).

Advances in the past decade have seen *E. coli* high-cell density fermentations producing high yields of human antibody fragments (Fabs) through a secretion mechanism and engineered intrinsic stability, allowing up to 40-fold increases in production over more commonplace periplasmic expression systems (Demarest et al., 2006, Rios, 2012). Whilst these advances keep microbial platforms such as *E. coli* competitive with mammalian platforms such as CHO, mammalian platforms remain optimised for situations where complex and intricate post translational protein modification is required. However synthetic biology has improved understanding of complex biological systems to the degree where such post translational modification is becoming possible in *E. coli* (Wang et al., 2001), and these strategies need further investigation and improvement.

1.6 Bioprocess monitoring

The bioprocess engineering community seeks to gain the optimum product yield and reproducible product quality from the host cell chassis with minimal resource expenditure, while shortening process development and time to market. To achieve these aims cellular production processes must be understood and modelled in order to extract the maximum production efficiency from the systems. Advancements to bioprocess monitoring technologies are providing increasingly sophisticated assays and platforms with which to gather the data necessary for the realisation of these aims (Clementschitsch and Bayer, 2006).

A desirable bioprocess monitoring assay is reliable whilst being relatively inexpensive and easy to perform. For these reasons assays can be improved in terms of their turnover time to gain a result, ease of use and reliability of results. Novel assays that

allow the detection of new factors are also desirable, for example, assays that are able to quantify the presence of plasmids within bacterial chassis (Lee et al., 2006a) without the use of antibiotics (Skulj et al., 2008).

1.6.1 Scope for improvement in Bioprocess monitoring

Regulatory frameworks such as Process Analytical Technology (PAT) are creating drivers for development of monitoring technologies (Kaiser et al., 2008). The objective of the PAT framework is to increase process understanding in order to support innovation and efficiency in the development of biopharmaceuticals. To achieve this, the integration of on-line and at-line monitoring systems is essential in order to gather as much information on bioprocesses as possible (Gnoth et al., 2007). Due the implementation of the PAT initiative, industrial facilities are strategically increasing their use of timely monitoring systems, or those with the capacity for in-line, at-line or on-line data collection (Table 2).

Table 2: Definitions of monitoring classes in bioprocessing

System	Definition	Example
In Line	The sample is measured within the process stream, and not removed. This can be invasive or non-invasive.	pH probe, DO2 probe
On-Line	The sample is automatically removed from the process stream and analysed by equipment usually built into the fermentation apparatus, then it might be returned or disposed of.	Mass Spectrometry
At-line	The sample is isolated from the process stream and analysed in situ by separate equipment. It is	Spectrophotometric measurement of optical density

	disposed of after.	

1.6.2 Biomass and Growth

Perhaps the most critical parameters to monitor during bioprocessing are the biomass, viability and growth kinetics of the production platform. Traditional off-line methods used to determine biomass include dry cell mass quantification through weighing. The desire for in-situ, on-line and at-line approaches to biomass quantification has seen the development of optical probes capable of online biomass determination (Kiviharju et al., 2008). Other approaches to biomass monitoring have been explored, such as those that expose cells to radiofrequency electrical fields to determine the proportion of viable cells with intact cell membranes (Soley et al., 2005).

1.6.3 Gasses

Monitoring of gasses, mainly oxygen and carbon dioxide, is important feature of bioprocess monitoring. Dissolved oxygen (DO) concentration is an important parameter to monitor and control as it directly impacts cellular metabolism and therefore production (Ozturk and Palsson, 1990). Data on DO is commonly gathered in-line, using a probe submerged in culture media, although on-line systems exist to measure off-gas also exist (Teixeira et al., 2009). The monitoring of dissolved carbon dioxide (DCO_2) is an important requirement for mammalian cell culture. CO_2 is used as pH control and maintaining it at a correct level (usually 5%) helps maintain optimum cell growth and secondary metabolite production (Miller et al., 1988).

1.6.4 Metabolites

Sugars, amino acids and other such metabolites are key nutrients for the support of cellular growth in fermentation, and as such are key targets for monitoring. By constructing models of metabolic profiles and how metabolites are utilised by candidate

bioprocess platforms, the bioprocess engineer can facilitate high throughput screens which identify the cellular phenotypes with the highest production capacity (Mapelli et al., 2008). Several approaches to metabolite monitoring have been described including monitoring of off-gas and using chemical-ionization mass spectrometry (CIMS) (Heinzle, 1992). The most recent approaches to metabolite monitoring are biosensors that are biological macromolecules, such as fluorescent reporter proteins (Constantinou and Polizzi, 2013).

1.6.5 Proteins and bio-products

A critical endeavour of bioprocess monitoring is the monitoring of product yield and quality and as such proteins and bioproducts are key bioprocess monitoring targets. Traditionally, western blotting has been used to identify and analyse target proteins (Shacter et al., 1994). Chromatography is also key technology with which to capture and analyse proteins, and methods such as high performance liquid chromatography help researchers study the effect of different fermentation parameters on final product quantity (Chen and Horváth, 1995). These techniques are however limited to offline use.

1.6.6 Gene expression

In order to be able to better understand and control production in host cell chassis a detailed understanding of the gene regulatory networks that govern metabolism and synthetic pathways are required. Transcriptome analysis can be conducted to discover how conditions within bioreactors affect protein synthesis. Cultivation within a bioreactor exposes host cells to a variety of environmental stresses, including nutrient gradients, oxygen deprivation and overproduction of recombinant proteins (Zhang et al., 2010). A detailed understanding of how these affect cell metabolism is extremely advantageous in process optimisation as it facilitates understanding of how cells adapt to these conditions, and can be used as a tool to build a library of biomarkers that signal when cells are experiencing conditions that will limit or halt production (Pioch et al., 2007). It also has the potential to influence host cell optimisation by engineering cells to better combat any deleterious effects (Wang et al., 2009).

Techniques for such analyses can include fusing gene promoters to reporter proteins such as GFP or luciferase (Weber et al., 2006). A major contributing technology to understanding gene expression is analysis using quantitative PCR (qPCR), especially as the library of complete genomic sequences for organisms is growing (Shendure and Ji, 2008). It is however largely limited to off line measurement.

1.7 Quantitative PCR as a bioprocess monitoring tool

qPCR has revolutionised the analysis and monitoring of nucleic acid and genetic function, and has found applications as a platform technology in areas as diverse as disease diagnostics (Planell-Saguer and Rodicio, 2011), oncogene detection(Kristensen et al., 2011) and transcriptome analysis (Morozova et al., 2009). As long as a robust assay is developed, qPCR has a number of diagnostic applications. The combination of ease of use, specificity, sensitivity and low risk of contamination has introduced real time PCR as a replacement for to traditional techniques such as culture-based or immunoassay-based diagnostics (Bravo and Procop, 2009). It is also fast becoming a transformative tool for monitoring bioprocess host systems (Lee et al., 2006b; Skulj et al., 2008) and has applications ranging from the tracking of host cells in fermentation to detecting and quantifying infection.

1.7.1 Sample preparation and target selection considerations for bioprocess monitoring by qPCR

Sample preparation is normally performed to remove inhibitors that might impact the accuracy and sensitivity of PCR-based assays (Dineva et al., 2007). Ionic detergents, phenol and metal salts that may be present in growth media can all inhibit PCR. Kits and reagents used to extract, purify and preserve DNA can also influence PCR and bring the risk of introducing error through loss of DNA (Miller et al., 1999), introduction of inhibitory biological material from disrupted cells and co-purification of chemical inhibitors (Schrader et al., 2012).

Commercial sample preparation kits, including membrane and bead-based systems, often involve protocols that comprise 20 steps or more and can take over one hour to

perform. This serves to significantly decrease assay throughput times and means that they are unsuitable for use at-line. Some DNA isolation kits have also been shown to produce false-positive results due to the presence of contaminant DNA in the kit, the level and source of which varies between manufacturer and batch. Queipo-Ortuño et al tested 7 commercially available DNA extraction kits for contamination and found that 6 of these had contamination from a range of organisms (Queipo-Ortuño et al., 2007).

When selecting targets for future study, it is important they have well-published and accessible sequences. Furthermore, if attempting to determine absolute copy number, all targets should be selected on the basis of them being single copy, to ensure only one is present on a genome and thus the quantification of target is directly proportional to the quantity of copy number.

1.7.2 Standardisation of bioprocess monitoring

The optimisation and scale up of bioprocesses and synthetic biology systems brings with it the challenge of standardising bioprocess analytics. A set of common metrics and standardised methods will allow for greater quality control and the capture of data sets comparable across multiple instruments and laboratories (Müller and Arndt, 2012). Once established, Synthetic Biology standards at scale could compliment bench scale standards that are already available (Kelly et al., 2009). Successful standardisation of bioprocess monitoring process is also likely to enable greater process understanding and help industrialists better demonstrate regulatory compliance (Kaiser et al., 2008).

1.8 Aims and objectives

At-line bioprocess monitoring by qPCR is currently unfeasible, due to lengthy sample preparation stages and an arduous workflow that limits assay throughput time. Furthermore a lack of standardisation has resulted in a diverse range of protocols, targets and primers, which limits reproducibility. The aim of this project is to therefore improve the workflow of qPCR for monitoring of biochemical engineering through a reduction in sample preparation steps and time. This is to increase the throughput of PCR assays, whilst maintaining reproducibility and accuracy, with a view to generating

the first viable, at-line and regulatory compliant qPCR bioprocess monitoring tool. Whilst we recognise that current sample preparation kits are routinely used to gain data, our hypothesis here is that they are not needed and serve to increase the throughput time to an extent that at-line analysis is rendered unachievable. A secondary aim is to also introduce a level of standardisation in order to increase reproducibility.

The main objectives of this project that will contribute to achieving the aim are as follows:

- We will identify single copy gene targets and develop a primer design and screening strategy, to ensure high efficiency PCR reactions are conducted and assays are compliant with widely adopted PCR regulatory guidelines (Bustin et al., 2009).
- We will explore increasing the how drastically increasing assay throughput time, by removing many sample preparation stages, affects e-pPCR accuracy across three model organisms
- We will measure the extent to which rapid sample preparation affects qPCR efficiency and quantitative accuracy
- For the first time this project will also introduce to introduce Synthetic Biology-compliant and industrially robust standards for bioprocess monitoring by PCR in order to increase reproducibility.
- Alternative methods of PCR data analysis will be explored and existing methodologies will be adapted to allow for standardisation, and a community wide proposal will be outlined for the adoption of standards by the qPCR and Synthetic Biology community.
- An assay will be developed that implements rapid sample preparation and synthetic biology standards, in order to monitor the evolution of plasmid copy numbers within an *E. coli* production strain, during the course of a fermentation
- All assays developed will be directly transferrable to rapid-cycle platforms, such as the BJS Biotechnology Xxpress thermocyclers, with the prospect of allowing for at-line analysis.

Chapter 2

2 Materials and Methods

All reagents were of molecular biology grade unless otherwise stated. All solutions were prepared using molecular biology grade water (Millipore, Billerica, USA).

2.1 Cell cultivation

2.1.1 *P.* pastoris cultivation

The production strain used was *P. pastoris* GS115 (Invitrogen) expressing a recombinant human protein. A single colony from a yeast-peptone-dextrose (YPD) agar plate (1% w/v yeast extract and peptone, dextrose, and agar all at 2% w/v) was used to inoculate 50mL buffered glycerol complex medium (BMGY) broth (1% w/v yeast extract, 2% w/v peptone, 100 mM potassium phosphate pH 6, 1.34% w/v yeast nitrogen without amino acids, 1% v/v glycerol, and 0.4µg/mL biotin) in a 250mL shake flask which was incubated at 30ºC, with 250 RPM agitation for 15 hours, after which typically OD_{600}=2 was reached. This inoculum was used for a working cell bank (WCB) of over twenty 1mL vials stored at -20ºC each containing 500µL inoculum mixed with 300µL 80% v/v glycerol.

For shake flask cultivation prior to fermentation, a WCB vial was thawed on ice and 100µL of the glycerol stock solution used to inoculate 100mL BMGY in a 500mL shake flask before incubation at 30ºC, with 250 RPM agitation until OD_{600} =50 was reached. 18mL of this inoculum was used to inoculate 550mL Basal Salt Media (BSM). BSM consisted of 26.7mL 85% w/v H_2PO_4, 0.93g $CaSO_4$, 18.2g K_2SO_4, 14.9g $MgSO_4.7H_2O$, 4.13g KOH, 40g glycerol and 12mL 'Pichia Trace Metal 1' (PTM1) solution (6.0 g/L $CuSO_45H_2O$, 0.08 g/L Nal, 3.0 g/L $MnSO_4.H_2O$, 0.2 g/L $Na_2MoO_4.2H_2O$, 0.02 g/L H_3BO_3, 0.5 g/L $CoCl_2$, 20.0 g/L $ZnCl_2$, 65.0 g/L $FeSO_4.7H_2O$, 0.2 g/L biotin, and 5.0 mL/L 96% H_2SO_4) per litre dH_2O. Bioreactor cultivation was performed according to a commercial protocol ("Pichia Fermentation Process Guidelines," 2002).

2.1.2 CHO cultivation

Glutamine synthase (GS) Chinese hamster ovary (CHO) cells stably expressing an immunoglobulin G4 (IgG4) were cultivated in T175 shake flasks (SF) according to the protocol described by Velez-Suberbie et al., (2013) until a viable cell count of 2x106 cell/mL was reached, determined by ViCell-XR cell viability analyser (Beckman Coulter, USA).

Bioreactor (BR) cultivation was performed in a 3L Applikon Appliflex (Applikon, Holland) flexible rocked bag bioreactor, controlled by an Applikon EZ controller system. Temperature was kept at 37°C with the dissolved oxygen (DO) set-point at 30% and the pH set-point at 7.1±0.05. The bioreactor was inoculated with at 2x106 viable cells per mL at mid-exponential phase and cultured for 14 days. Glucose concentration was maintained at 150g/L, as determined by NOVA Bioanalyser 400 (NOVA Biomedical, Waltham, USA), by supplementing with 10x concentrated CD-CHO media (Life Technologies, Paisley, UK).

2.1.3 *E. coli* cultivation

E. coli W3110 (K12) was transformed with pTODD A33 plasmid encoding an antibody fragment (Fab'). pTODD A33 was kindly donated by UCB Celltech.

200mL of defined media (10g tryptone, 5g yeast extract, 5g NaCl per 1L dH_2O), containing 10µg mL^{-1} tetracycline, was used as a starter culture and cells were cultivated in shake flask, at 37°C and 200RPM. Upon cell density reaching an OD_{600} of 2.5, this starter culture was used as 10% (v/v) inoculum for defined media (5.2g $(NH_4)_2SO_4$, 4.4g $NaH_2PO_4.H_2O$, 4g KCl, 1.04g $MgSO_4.7H_2O$, 10mL SM6e trace elements (100g $C_6H_8O_7$, 5g $CaCl_2.6H_2O$, 2.46g $ZnSO_4.7H_2O$, 2g $MnSO_4.4H_2O$, 0.5g $CuSO_4.5H_2O$, 0.427g $CoSO_4.7H_2O$, 9.67g $FeCl_3.6H_2O$, 0.03g H_3BO_3, 0.024 $NaMoO_4.2H_2O$ per 1L dH_2O), 4.2g $C_6H_8O_7.H_2O$, 141g glycerol, 0.25g $CaCl_2.H_2O$ per 1L dH_2O) shake flask fermentation, grown at 30°C and 200RPM.

Upon defined media culture reaching an OD_{600} of 5, 10% (v/v) was used to inoculate defined media in a New Brunswick 7L Bioreactor (Eppendorf, Hamburg, Germany). A fed-batch protocol was then used that has been described previously by Tustian et al. (2007).

2.2 Sample preparation

Following sample preparation, aliquots were made and immediately stored at -20C until further analysis.

2.2.1 Nucleic acid purification

Shake flask and bioreactor samples were centrifuged at 10,000 RPM for 3 minutes, re-suspended in an equal volume of cell lysis buffer (2% Triton-X100, 1% SDS, 100mM NaCl, 10mM Tris-HCl, 1mM EDTA) and freeze-thawed twice by incubating at -80ºC for 3 minutes and 95ºC for 1 minute. Nucleic acid was extracted using standard phenol/ethanol extraction (Wilson, 2001) and purified DNA was resuspended in equal volume of 10mM Tris (pH 7.5) which was then split into aliquots of equal volume and stored at -20ºC. A given aliquot was thawed once for experimentation and any unused portion of the aliquot discarded.

2.2.2 Sonication

400µL shake flask and bioreactor samples were centrifuged and re-suspended in dH_2O to a total volume of 400µL. A Soniprep 150 sonicator (MSE, London, UK) was used to subject samples to a 10 second cycle of 100% amplitude sonication, followed by 10 seconds rest, three times. Total procedure duration was 5±2 minutes (n=20).

2.2.3 Heat lysis

An Eppendorf Thermomixer (Eppendorf, Hamburg, Germany) was used to incubate at 95°C for 10 minutes.

2.3 Oligonucleotide design

2.3.1 Primers

Experimental primer sequences (Table 1) were designed in accordance with MIQE guidelines [14] and also refined for specificity with the National Center for Biotechnology Information (NCBI) 'Primer-blast' tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/, accessed 23.11.14) and for self-annealing with the NCBI 'PCR primer stats' tool (http://www.bioinformatics.org/sms2/pcr_primer_stats.html, accessed 23.11.14).

2.3.2 Cal 1 sequence

Cal1 (Rutledge and Stewart, 2010) LRE-qPCR primer sequences were used for calibration of data analysed by linear regression.

2.3.3 Plasmid design

A plasmid was designed that contained 300bp of a 16s RNA signature gene found in the 5 species of *Mycoplasma* that are commonly found in infected mammalian cell culture (Kong et al., 2001). The gene was inserted into a pUC57 plasmid by Eurofins MWG Operon (Acton, UK) and propagated using standard molecular biology techniques.

2.4 PCR conditions

2.4.1 e-pPCR

Reactions were carried out in a total volume of 50µL, with 5µL of 10x $MgCl_2$ polymerase buffer (100mM Tris/HCl, 15mM $MgCl_2$, 500mM KCl), 0.5µL Taq polymerase, 1µL 50µM dNTP (Sigma Aldrich, St. Louis, MO, USA), 5µL of material containing template DNA and 2.5µl of primer at a concentration of 1µM (to give a final concentration of 500nM of each primer per reaction). A Veriti 96 well thermocycler (Applied Biosystems Grand Island, NY, USA) was used with a cover heated to 105ºC. Each PCR was run for 40 cycles of: 95°C for 5 seconds, 57°C for 5 seconds, 72°C for 30 seconds.

2.4.2 qPCR

Reactions were carried out in a total volume of 20µL, with each reaction containing 10µL of 2x SsoAdvanced SYBR Green Supermix (BioRad, Hercules, CA, USA), 5µL of material containing template DNA and 1µl of primer at a concentration of 1µM (to give a final concentration of 500nM of each primer per reaction). Reactions were performed in a CFX Connect Real-time PCR Detection System (Bio-Rad, Hercules, CA, USA) with a cover heated to 105ºC. Each reaction was run at a total of 40 cycles, with the same cycling conditions as above.

2.4.3 PCR efficiency estimation

Efficiency of qPCR reaction was calculated from a standard curve constructed from purified DNA samples that were serially diluted from neat stock and amplified by qPCR in parallel. Cq values from amplification curves were generated by the fit-point method and plotted against log dilutions. A straight line was drawn between data points corresponding to a coefficient of determination (R^2) of 0.99. Linear regression was then applied to calculate efficiency (*E*), with the equation (Pfaffl, 2001):

$$E = 10^{(\frac{-1}{slope})}$$

2.5 Nucleic acid quantitation

2.5.1 Spectrophotometry

DNA concentration was determined by spectrophotometry, using a Nanodrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). The molar absorption coefficient of DNA and sample dilution factor were used to infer DNA concentration (Efiok et al., 2000). Shake flask process streams were analysed undiluted, and diluted 10 and 100 fold, and DNA mass at higher dilutions was then extrapolated from these analyses. When analysing bioreactor process streams, it was found that DNA mass of samples at OD_{600}=800 could not be accurately quantified, so 10, 100 and 1000 fold dilutions were analysed (Figure 4).

2.5.2 Densitometry

Samples were run on 1% agarose gels stained with ethidium bromide and bands visualised using a GelDoc 2000 device (BioRad, Hercules, CA, USA) and Quantity One software version 4.6.8 (BioRad, Hercules, CA, USA). ImageJ software (version 1.46r, National Institutes of Health, Bethesda, USA) was used to select a region of the gel image, either a lane or a band, containing a known mass of DNA and a brightness value captured. A selected region of the same size and shape was then used to capture brightness of a region of unknown DNA concentration on the same gel. Background noise was subtracted using the 'Background Subtract' function provided by the ImageJ software (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/, 1997-2014).

2.5.3 Standard curve

The standard curve generated as described previously was used to estimate copies of target in samples contaminated by cell debris. Cq values of contaminated samples were plotted along the standard curve and converted into copy number using the equation:

$$target\ copy\ number = 10^{\left(\frac{Cq-b}{m}\right)}$$

Where *b* is the y-intercept and *m* is the slope of the standard curve.

2.5.4 LRE

LRE PCR, as described by Rutledge (Rutledge, 2008), was also applied to estimate copy numbers. LRE analyser v. 0.97 (Rutledge, 2011) was used according to developers instructions. Purified samples, with known DNA mass, were used to calibrate the program and provide information on copy number in samples contaminated with cell debris.

2.5.5 Cy0

The Cy0 was calculated by uploading raw fluorescent data to the Cy0 portal, found at www.cy0method.org.

2.6 Flow Cytometry

Flow cytometry was performed using an Accuri C6 flow cytometer (BD Biosciences, USA), using bis-oxonol (BOX) and propidium iodide (PI) stains according to previously described methods (Lewis, et al., 2004). Samples from lab-scale bioreactor fermentation were diluted to an OD_{600} of 1.0 AU and 10 µL was combined with 990 µL of staining solution (50 µg/mL BOX, 40mM EDTA, 20 µg/ml PI, in PBS) and incubated at room temperature for 8 minutes. Viable cells/mL was calculated by counting cells remaining unstained, or stained with BOX.

2.7 HPLC

Volumetric Fab' concentration was assayed using Agligent 1200 (Agligent Technologies UK Ltd, West Lothian, UK), calibrated using standards of pure Fab' at known concentration. A protein G Hi-Trap column (GE Healthcare, Uppsala, Sweden) of capacity 1mL was used. The samples were filtered using 0.22µm syringe filters into appropriate HPLC vials for measurement. Adsorption was performed using 20nM phosphate buffer at pH 7.4 and the product was eluted using 20mM phosphate buffer at pH 2.5. The amount of fab eluted was measured by recording absorbance at 200nm.

Chapter 3

3 Assay optimisation through primer screening and validation

3.1 Introduction

The design and validation of suitable PCR primers is an important consideration in PCR assay optimisation. Poorly designed primers can reduce or prevent amplification through lack of specificity, or produce unspecific products through dimer formation. Dimer formation can occur when sequences within the primer share similarity in nucleotide sequence such that the primers will self-anneal and form dsDNA without the presence of template (Figure 3-1). The formation of primer dimers has the potential of confounding analysis as they can generate an artificial fluorescent signal, resulting in overestimation of product or false positive detection (Hyndman and Mitsuhashi, 2003).

To ensure specificity and to minimise the potential of dimer formation, a number of bioinformatics tools are available that calculate primer sequences according to user-selected criteria. The user inputs the target DNA sequence and selects parameters such as primer length, primer melting temperature and product size and suitable candidates are generated. Candidate primer sequences can then be further screened *in-silico* to determine the potential for primers to amplify non-specific sections of the target genome and to self-anneal and forming dimers, ensuring the final primer pairs generated have a low chance of forming secondary products.

Once primers are obtained their suitability can be confirmed by amplifying the target and analysing the amplicons. If SYBR-Green detection chemistry is used, specificity can be confirmed by melt-curve analysis, where amplicons are subject to a steady increase of temperature and change in fluorescence is monitored. As specific primers will form amplicons of uniform length they will dissociate at the same temperature, causing a rapid fluctuation in fluorescence. Different detection chemistries can be analysed by gel-electrophoresis to ensure product size is consistent with expectations.

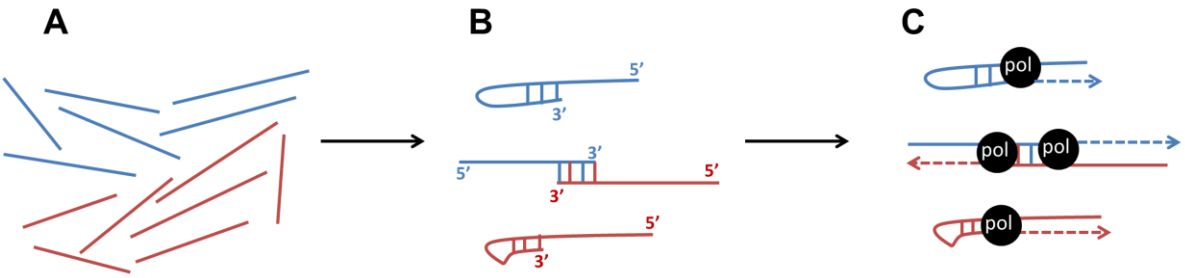

Figure 3-1 The need for *in silico* screening of primer designs. Primer dimer formation and amplification of non-specific products result when two sets of primers free in reaction (A) bind to either themselves or each other (B) and as a result polymerase will bind to the newly formed dsDNA section to synthesise a non-specific product (C).

Here we aim to develop a design strategy that will produce optimised primers with no dimer formation, which are specific to the range of templates used throughout future studies and will generate amplicons according to a high efficiency. We also aim here to directly compare the SC qPCR and LRE-qPCR methods on pure chromosomal and plasmid targets.

3.2 Results

3.2.1 Design strategy

Ten candidate primer sequences were generated for each target according to the design criteria outlined in Table 3, using the NCBI primer blast database. Primer length was chosen to ensure maximum specificity to the target whilst minimising the chance of the primer self-annealing. Smaller amplicons can be synthesised with a higher efficiency, thus amplicon size was kept below 250bp. Primer melting temperatures were determined by the nearest neighbour method and were chosen based on the cycling regimen used in future studies. If the target is non-genomic DNA, specific to this exogenous DNA over the genomic sequence must also be determined at this point, to prevent the chance of amplifying from the wrong target. To achieve this any potential primer sequences can be screened against the genomic sequence of the host cell species.

From these 10 candidates, three were selected for *in vitro* screening based on their low chance of forming dimers, following screening with a bioinformatics program (Vallone and Butler, 2004). Testing was conducted over two biological replicates, with each being analysed over three technical replicates. Primers concentration was set at 500nM per reaction and each candidate pair was used to amplify a dilution series of the target. Primers were selected based on those that could distinguish concentration difference over the largest dynamic range, produced the lowest overall Cq values and highest efficiency when analysed by LRE-qPCR, and produced only a single dissociation peak when subjected to melt-curve analysis.

Table 3 Design criteria used in generating primer pairs

Criteria	Value
Amplicon size	60-250bp
Primer length	19-23 nucleotides
Minimum melting temperature	50ºC
Optimum melting temperature	68ºC
Maximum melting temperature	60ºC
GC content	35-65%

3.2.2 Transketolase (Tkt)

Transketolase is a single copy gene coding for an industrially relevant enzyme and it's sequence is published in the literature (De Schutter et al., 2009). The gene is found at the genomic locus PAS_chr1-4_0150, and here it is amplified from purified *P. pastoris* gDNA. Primer set 2 shows consistently lower Cq values in comparison to the two other primer pairs and an associated low amplification efficiency. As all three show no dimer formation when subject to melt-curve analysis (Figure 3-2), primer 2 was chosen as the optimal primer pair and carried forward into subsequent experiments.

Table 4 Candidate sequences generated for Tkt

Primer set	5' Sequence	3' Sequence	Amplicon size
1	GCCTGTTGCCTCAAGAAAGC	AAGTTGGCACCATAAGCGGA	243

2	CTCCACTGAAGCCTGATGAC	ACTCCTGACCCAAATCTGGA	104
3	TCCGCTTATGGTGCCAACTT	GGCGTAGGACACGAAGTTCA	63

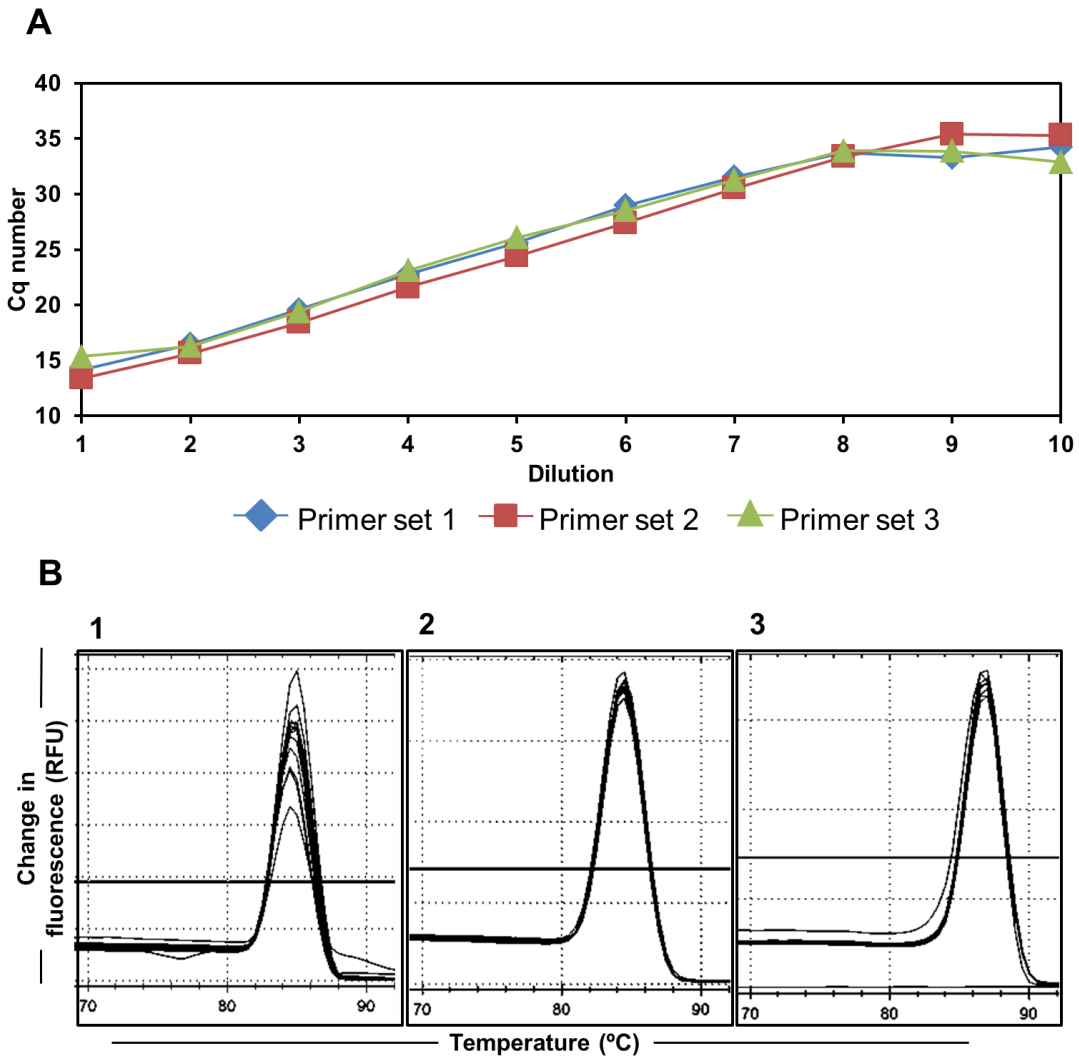

Figure 3-2 In vitro screening of Tkt primer candidates

3.2.3 Glyceraldehyde 3-phosphate dehydrogenase (GapDH)

GapDH is a single copy gene found in mammalian cells (Gene ID: 100736557) and here it is being amplified in purified Chinese Hamster Ovary (CHO) gDNA. It's sequence is reported in the literature as a qPCR housekeeping gene (Sikand et al., 2012). Of the three sets of primers designed, primer set three showed consistently low Cq values and the highest overall amplification efficiency. As melting curve analysis indicated that no primer pair forms dimers (Figure 3-3), primer set 3 was chosen as the optimal primer pair and taken forward in future studies.

Table 5 Candidate sequences generated for GapDH

Primer	5' Sequence	3' Sequence	Amplicon size
1	CTGACATGTCGCCTGGAGAA	AGGTGACTACTCAAGCCCCA	86
2	TTGGGGCTTGAGTAGTCACC	CCCCAGCATCAAAGGTGGAA	179
3	CATCACCATCTTCCAGGAGC	CTTGGTTCACACCCATCACA	194

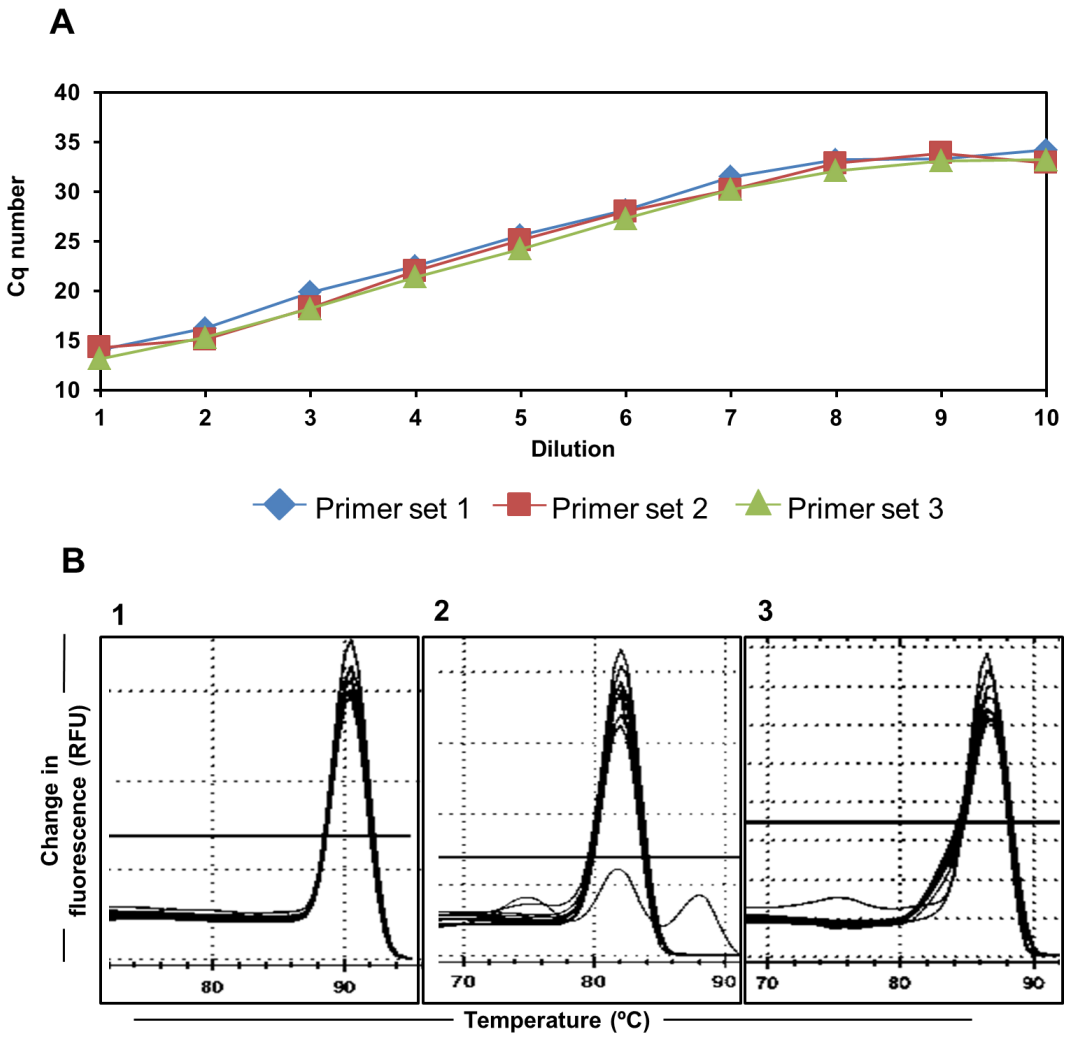

Figure 3-3 In vitro screening of GapDH primer candidates

3.2.4 Biotin retention locus (BirA)

BirA is a single copy gene found in strains of *E. coli* that encodes the enzyme Biotin Ligase (Gene ID: 12934397). It is a single copy target with a published sequence with industrial applications (Entemadzadeh et al., 2015). Of the three primer sets designed, primer sets 1 and 3 showed similar Cq values across the range of dilutions tested. Primer set one continued to produce significantly different Cq values between dilutions 8 and 9, indicating a higher dynamic range. Melting curve analysis shows primer set 3 produces non-specific products, indicated by the presence of a second peak. Therefore, primer set 1 was chosen as the optimal primer pair.

Table 6 Candidate sequences generated for BirA

Primer set	5' Sequence	3' Sequence	Amplicon size
1	ATCCACCCCTGATTAACGAC	CGGAAGTATTACGCAAGCTG	199
2	CAGGCAGGCTGTATCCTTTA	AAACACATTCAGACACTGCG	73
3	CGATCCTGCAGATAGAGGTC	ATCGTGATGGCGGAAGTATT	83

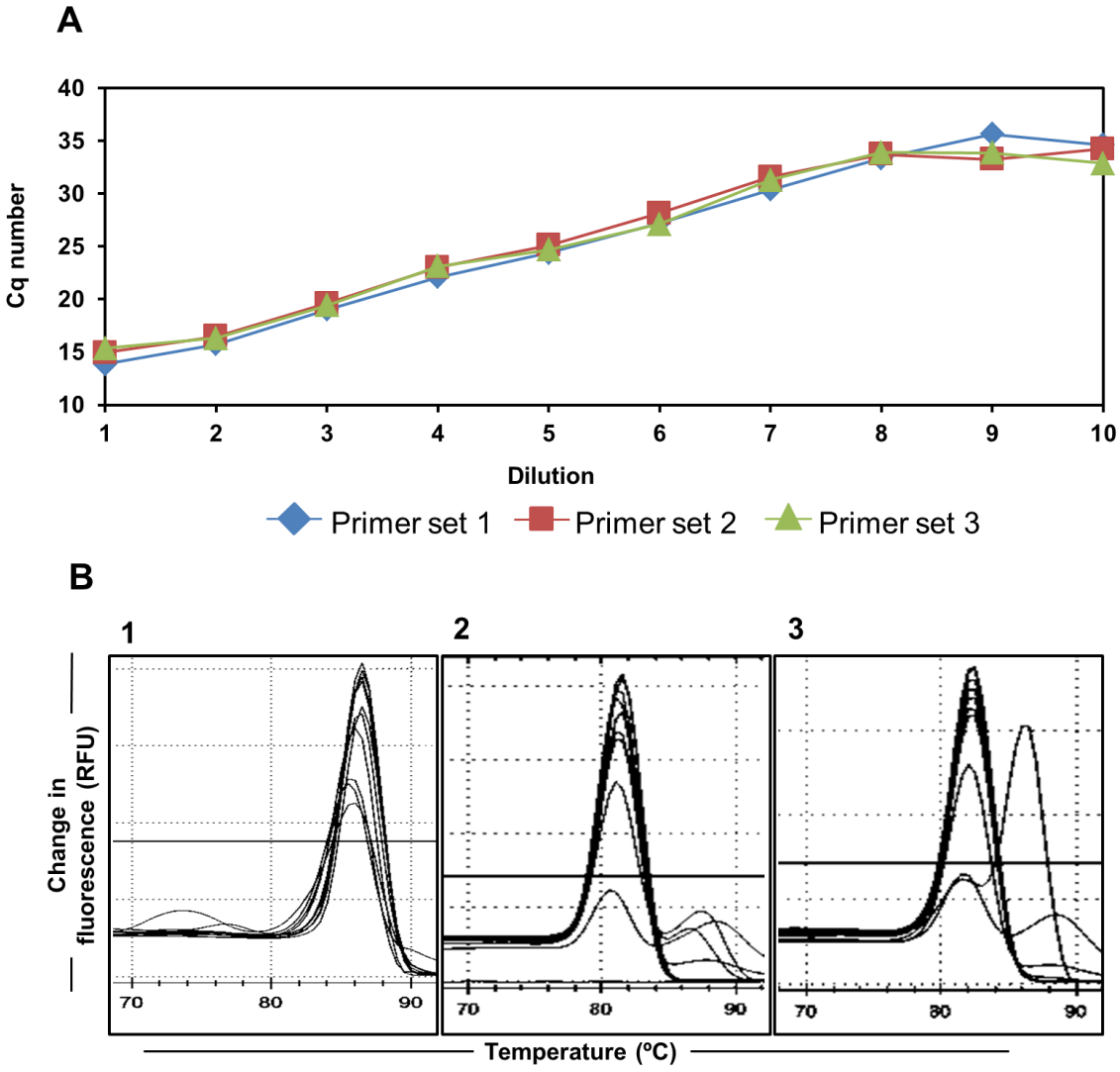

Figure 3-4 In vitro screening of BirA primer candidates

3.2.5 Fragment antigen binding (Fab) antibody fragment

A gene encoding a Fab antibody fragment contained within plasmid pTODD A33 was amplified by the three primer pairs. This gene is directly relevant to the plasmids function, with a known sequence and thus represents a good target. From this, primer set 2 shows consistently lower Cq values and a higher overall amplification efficiency. There is no evidence of dimer formation amongst any of the primer pairs or lack of specificity to the target, thus primer set 2 was chosen as the optimal primer set.

Table 7 Candidate sequences generated for Fab

Primer set	5' Sequence	3' Sequence	Amplicon size
1	ATGGCCTTCCGTGTTCCTAC	TGACTACTCAAGCCCCAACC	125
2	CCAGGCATCAAATTAAGCAGA	AAAGGGAATAAGGGCGACAC	195
3	AGAACATCACCCCTGCATCC	GGTGACTACTCAAGCCCCAA	207

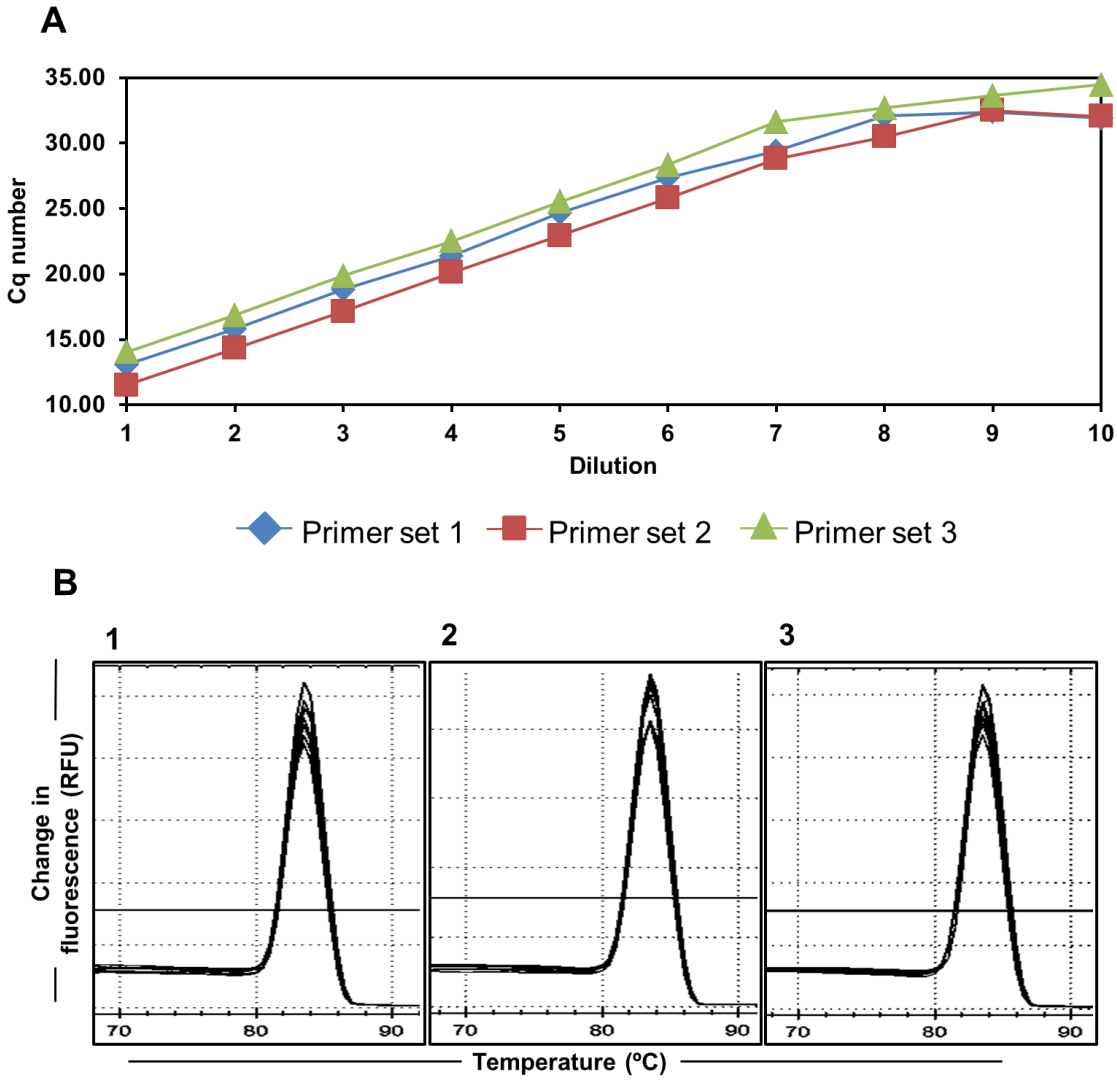

Figure 3-5 In vitro screening of Fab' primer candidates

3.2.6 Green fluorescence protein (GFP)

A gene encoding a green fluorescence protein was amplified from within plasmid pJDTI. The gene encoding the protein target is directly relevant to the plasmids function, with a recognised sequence and thus represents a good target. Whilst all three primer sets show comparable Cq values, set 1 is shown to differentiate between dilutions 7 and 8, indicating the possibility of this primer pair amplifying over a larger dynamic range. When subject to melt-curve analysis however, the formation of multiple peaks from primer pairs 1 and 3 indicate dimer formation, therefore primer set 2 was chosen for further experimentation.

Table 8 Candidate sequences generated for GFP

Primer set	5' Sequence	3' Sequence	Amplicon size
1	TTTCACTGGAGTTGTCCCAA	CCGTATGTTGCATCACCTTC	99
2	GAAGCGTTCAACTAGCAGAC	GGTCTCTCTTTTCGTTGGGA	129
3	TTTTCACTGGAGTTGTCCCA	TCCGTATGTTGCATCACCTT	101

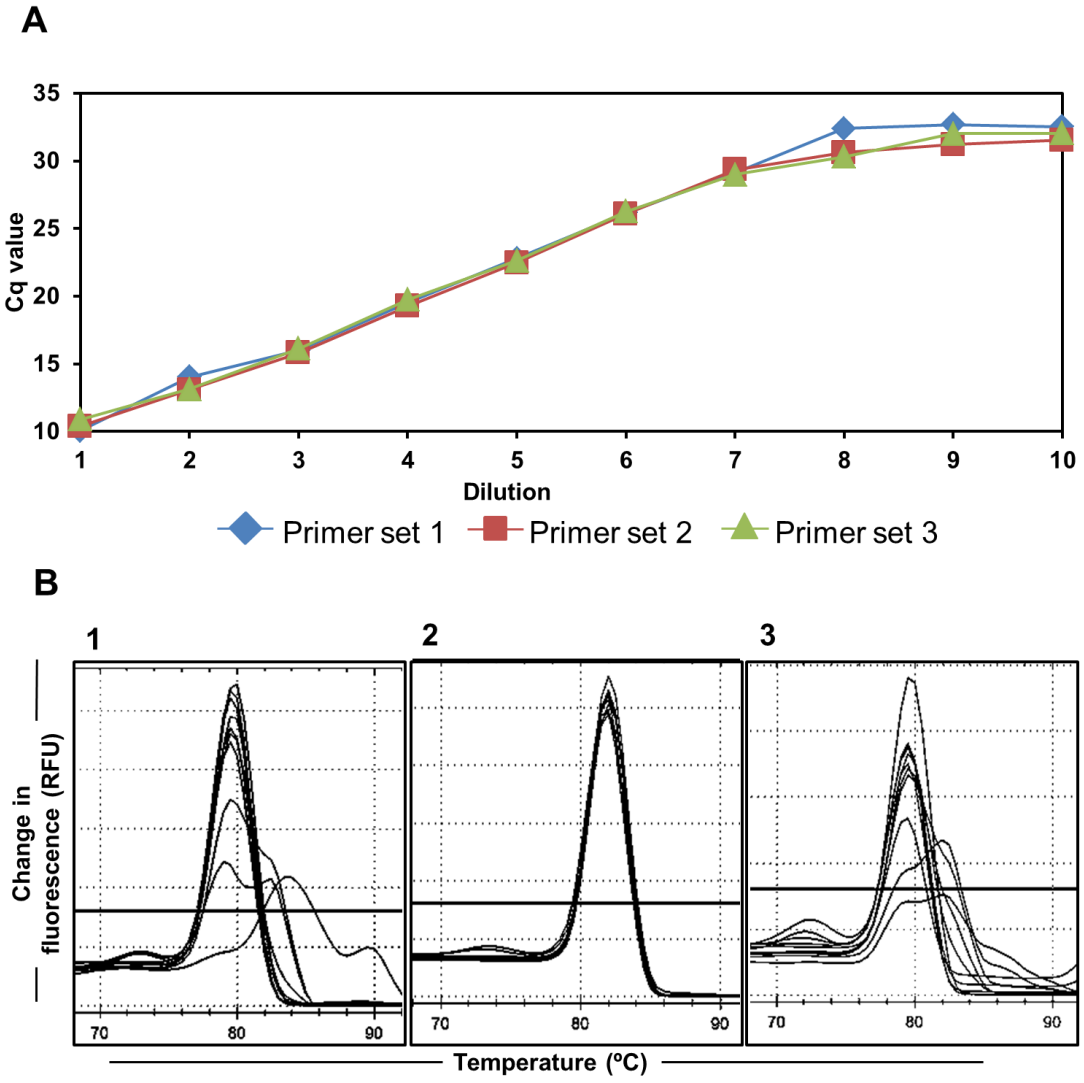

Figure 3-6 In vitro screening of GFP primer candidates

3.2.7 Mycoplasma consensus sequence

We designed a plasmid, pPROX2, encoding 300bp of mycoplasma DNA (Table 9) as a safe alternative to using live or attenuated mycoplasma. Whilst there are 20 species of mycoplasma known to infect mammalian cell culture, five species are identified in most infections. These are *M. arginine, M. fermentans, M. hyorhinis, M. orale* and *A. laidlawii*. The 300bp sequence we chose to use is conserved across five of the six mycoplasmas common to 90-95% of mammalian cell culture infections (Kong et al., 2001). Primer candidates were screened against the *Cricetulus griseus* (CHO) genome to ensure specificity. Figure 3-7 shows that, whilst primer sets 1 and 3 do not show dimer formation, primer set one has sensitivity over a wider dynamic range and so this set was taken forward for further study.

Table 9 300bp sequence inserted into plasmid

***Mycoplasma spp.* 16s RNA sequence**
TTGTACTCCGTAGAAAGGAGGTGATCCATCCCCACGTTCTCGTAGGGATACCTTGTTCGACT TAACCCCAGTCACCAGTCCTGCCTTAGGCAGTTTGTTTATAAACCGACTTCGGGCATTACCA GCTCCCATGGTTTGACGGGCGGTGTGTACAAGACCCGAGAACGTATTCACCGTAGCGTAGC TGATCTACGATTACTAGCGATTCCGACTTCATGTAGTCGAGTTGCAGACTACAATCCGAACT GAGACCGGTTTTTTGAGGTTTGCTCCATGTCACCACTTCGCTTCTCTTTGTA

Primer set	**5' Sequence**	**3' Sequence**	**Amplicon size**
1	AAACCGACTTCGGGCATTAC	GAAGTGGTGACATGGAGCAA	184
2	CCATCCCCACGTTCTCG	ACCATGGGAGCTGGTAATGC	113
3	TCCATCCCCACGTTCTCGTA	GAGCTGGTAATGCCCGAAGT	104

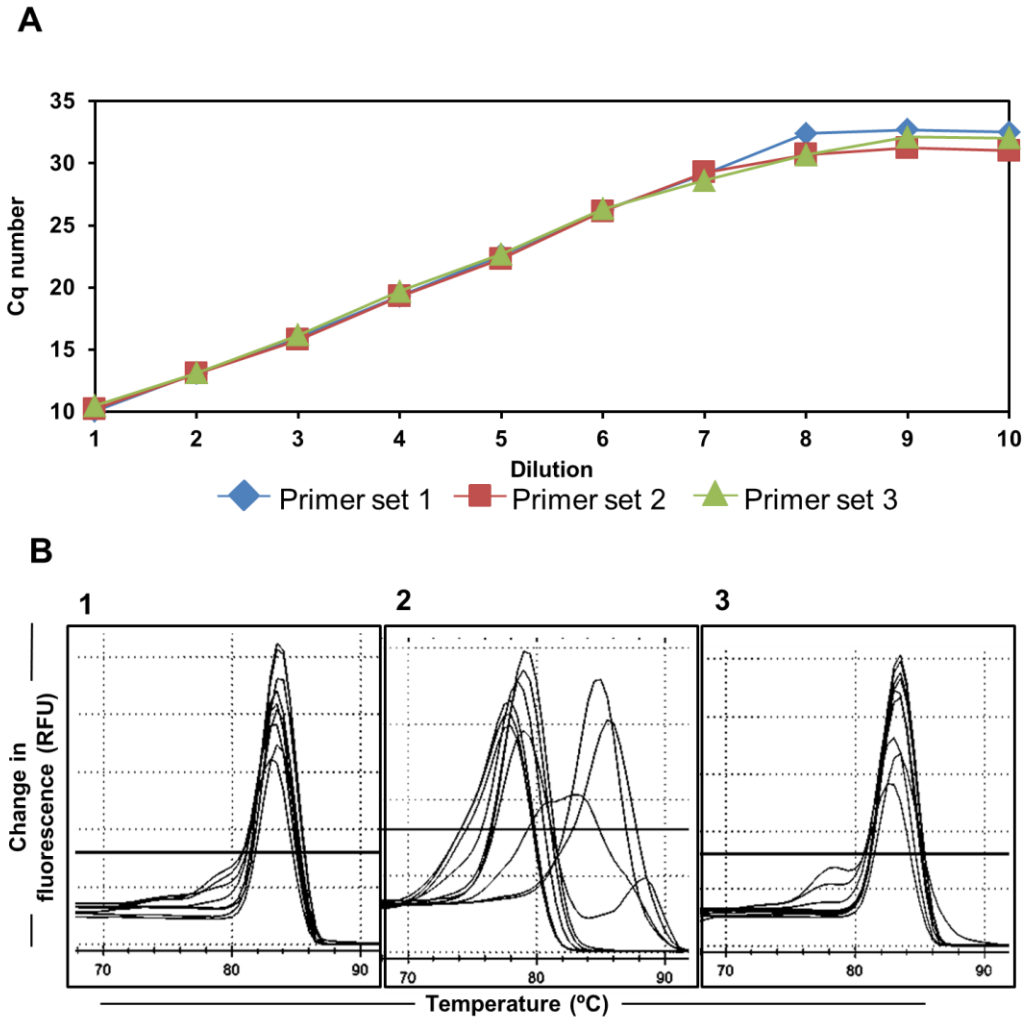

Figure 3-7 In vitro screening of mycoplasma consesus sequence primer candidates

3.2.8 Direct comparison of Standard Curve-derived copy number and LRE-qPCR derived copy number

During the testing phase of our primers, we examined each primer candidate against a purified target background. This should be close to ideal reaction conditions, as the reaction media should have no inhibitors present. We therefore took this opportunity to directly compare the quantitative ability of Standard Curve (SC) qPCR and LRE-qPCR, two data analysis methods used in future studies, on these inhibitor-free purified samples. This will allow us to assess whether the methods are comparable in ideal conditions.

Targets used in future studies will be present in either a chromosome or a plasmid. To qaccount for this we amplified the BirA *E. coli* chromosomal target and the Fab target of the pTODD A33 plasmid, in order to provide a comparison of the quantitative abilities of each method on both nucleic acid sources.

As can be seen from Figure 3-8 the profiles of SC qPCR and LRE qPCR match across the range of dilutions studied. This suggests that each method has the same quantitative capacity and means that each is suitable for assessment of the impact of cellular debris from rapid sample preparation that will be performed in future chapters.

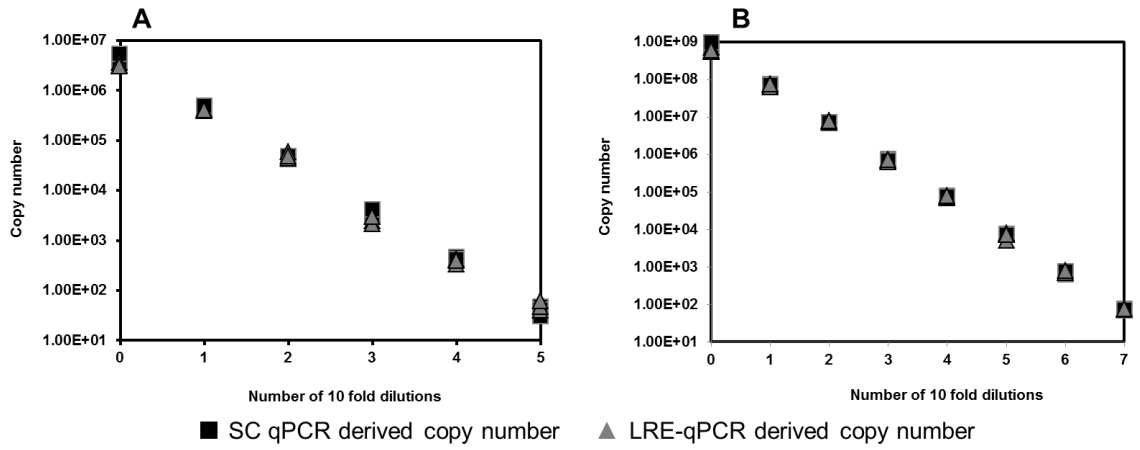

Figure 3-8 Direct comparison of quantitation by Standard Curve qPCR and LRE-qPCR. A is a direct comparison of SC qPCR and LRE-qPCR quantitation on the chromosomal BirA target on purified E. coli gDNA. B is a direct comparison of SC qPCR and LRE-qPCR on the plasmid target pTODD A33.

3.3 Discussion

The design strategy shown here, consisting of screening multiple primer candidates *in silico* for sequence homology and potential for non-specific amplification, followed by *in vitro* screening of three candidates for dimer formation, has ensured the generation and experimental use of primers specific to the target gene. By detailing the sequences used for future experiments and the methodology used to generate them, this optimisation strategy falls in line with that outlined by initiatives such as the Minimum Information for publication of Quantitative Realtime PCR Experiments (MIQE) guidelines (Bustin et al., 2009). This code of practice for how to perform and interpret qPCR studies facilitates reproduction of data by peers, and adherence to this ensures that data gathered can be experimentally validated.

Furthermore we have directly compared two methods of analysing the data, SC qPCR and LRE-qPCR. The results we obtained from this analysis suggest that the quantitative power of each data analysis method is comparable, when the reaction media is relatively free from inhibitors.

Table 10 Selected optimal primers

Organism, target	**5' sequence**	**3' Sequence**
CHO, GapDH	CATCACCATCTTCCAGGAGC	CTTGGTTCACACCCATCACA
P. pastoris, TKT	CTCCACTGAAGCCTGATGAC	ACTCCTGACCCAAATCTGGA
E. coli, BirA	ATCCACCCCTGATTAACGAC	CGGAAGTATTACGCAAGCTG
pTODD A33, Fab'	CCAGGCATCAAATTAAGCAGA	AAAGGGAATAAGGGCGACAC
pJDTI, GFP	GAAGCGTTCAACTAGCAGAC	GGTCTCTCTTTTCGTTGGGA

Myco	AAACCGACTTCGGGCATTAC	GAAGTGGTGACATGGAGCAA

3.4 Conclusions

- We have developed a screening method for the optimisation of PCR assays through primer design to specific criteria
- We screened 10 primer candidates for the possibility of forming dimers and specificity to the target *in silico* to generate three candidates for *in vitro* screening
- We screened three candidates *in vitro* for the possibility of forming dimers, as determined by a melt curve assay, and sensitivity over a 10 fold dynamic range
- The lead primer candidate taken forward was the primer that did not form dimers or has the highest sensitivity over the range studied
- We now have a set of optimised primers for further study, ensuring assays do not produce non-specific products and amplify according to a high efficiency
- The screening methods shown here are compliant with industry guidelines such as MIQE
- We directly compared the quantitative abilities of SC qPCR and LRE-qPCR, finding both to be comparable when conducted on a purified target

Chapter 4

4 Influence of high cell density *Pichia pastoris* cells on performance of PCR as a synthetic biology tool for bioprocess monitoring and contaminant detection

4.1 Introduction

The methylotropic yeast *Pichia pastoris* (*P. pastoris*) is often exploited industrially through the insertion of transgenes into the host genome (Inan et al., 2007). Synthetic biology has facilitated an increase in the level of genetic modification achievable and even enabled construction *in vivo* assembly of entire *Saccharmoyces cerevisiae* chromosome arms (Dymond et al., 2011) and chromosomes (Annaluru et al., 2014). Through elegant synthetic chromosome design these segments are able to undergo self-directed recombinative rearrangement to generate genotypically diverse populations. Disappearance or appearance of given sequences due to such rearrangements can, to an extent, be monitored using PCR-based methods. It is also critical for industrial application of yeast synthetic biology that such methods are robust to industrial settings. Industrial application of yeasts frequently involves cultivation of cells to high cell density (HCD), placing cells under extreme recombinant protein synthesis burdens or tolerance of high concentrations of small molecule substrates or products. Such environments have a strong potential to exert selective pressure on cells to inactivate transgenes or synthetic genes by gene mutation, loss or rearrangement.

In addition to actual scale environments, high throughput, microscale screening methods are increasingly being used to isolate biological variants and conditions best suited to industrial application (Baboo et al., 2012). Such microscale approaches ideally mimic industrial conditions with respect to factors such as cell density. Monitoring synthetic yeast genomes during industrial scale processes could reveal any possible effects of selection pressure or locus instability exerted by a given production step or set of conditions. However, current options for locus quantitation involve relatively lengthy approaches such as Southern blotting or preparation of samples for qPCR in which

DNA has been purified (Abad et al., 2010). These approaches tend not to be sufficiently rapid to enable at-line monitoring of yeast cells within scale industrial processes or their scaled-down mimics.

Taking this into account, it would be advantageous to develop PCR-based methods to quantify the abundance of sequence-specific DNA within yeast genomes, ideally with an absolute measurement standard and in a sufficiently rapid manner to enable at-line monitoring during an industrial process (actual or mimicked).

In this chapter we firstly test the hypothesis that the presence of non-DNA cellular *P. pastoris* material does not always compromise PCR accuracy. Specifically we are concerned to test if lengthy sample preparation procedures are necessary for accurate PCR-based assays of *P. pastoris* cellular material. We measure the influence of cellular material on the sensitivity of e-pPCR when used to detect the presence of a specific genomic sequence in a sample. We also measure the influence of *P. pastoris* cellular material on performance of qPCR for absolute target quantitation (Rutledge and Côté, 2003). Secondly, we test the hypothesis that the concentration and provenance of non-DNA cellular material are factors that influence PCR accuracy. To test this we perform PCR in the presence and absence of disrupted cells derived from high cell density cultivation in bioreactors and lower cell density cultivation in shake flasks. Finally, we test that hypothesis that LRE-qPCR (Robert G Rutledge, 2008) can match the accuracy of SC qPCR in the above conditions. We also discuss the suitability of LRE-qPCR as a synthetic biology standard. From the information gained through these experiments we address the timescales required to perform PCR by quantifying the extent to which sample processing is actually necessary to obtain accurate end point PCR (e-pPCR) and qPCR data from high cell density *P. pastoris*.

4.2 Results

4.2.1 Cultivation of P. pastoris

We cultivated a *P. pastoris* recombinant protein production strain in complex medium in shake flasks and took a sample when OD_{600}=50 for use as PCR template material

(Figure 4-1 A). This optical density is typical of both the early stages of industrial scale cultivation and the dilution steps that can be necessary for efficient dewatering of high cell density *P. pastoris* cultures (Lopes and Keshavarz-Moore, 2012). Shake flask material was used to inoculate a 550mL of BSM media in an Infors Multifors 1L bioreactor. When OD_{600}=800 was achieved, typically after 60-70 hours of methanol-induction (Figure 4-1B), a further sample was taken for PCR analysis. This represents a stage of cultivation when it is critical that recombinant protein yield and quality objectives have been met prior to harvest.

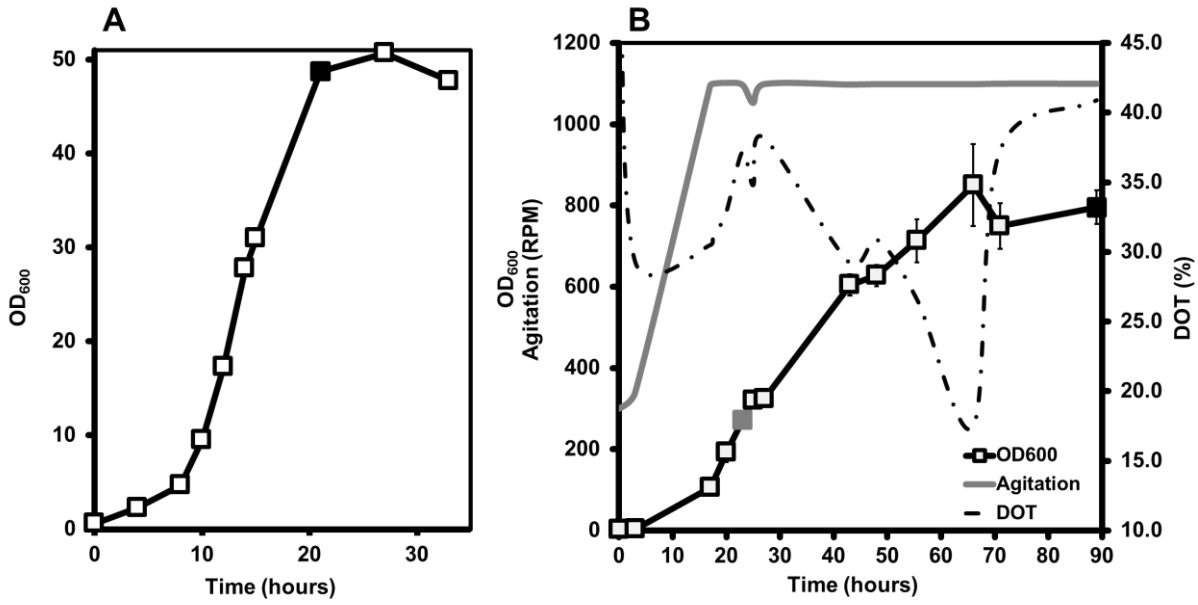

Figure 4-1 Shake flask and high cell density cultivation of P. pastoris production strain GS115. A methanol-inducible GS115 P. pastoris production strain was used to inoculate 100mL BMGY media in a 0.5L shake-flask (A) and an uninduced sample taken of cells entering stationary phase growth (filled square). 18mL of shake flask culture was then used to inoculate 550mL BSM media in an Infors Multifors 1L bioreactor (B). 24 hours post-inoculation methanol was added for induction of transgene expression and a sample for PCR analysis was taken 66 hours post-induction (filled square). Data sets are representative of n=3 experimental repeats.

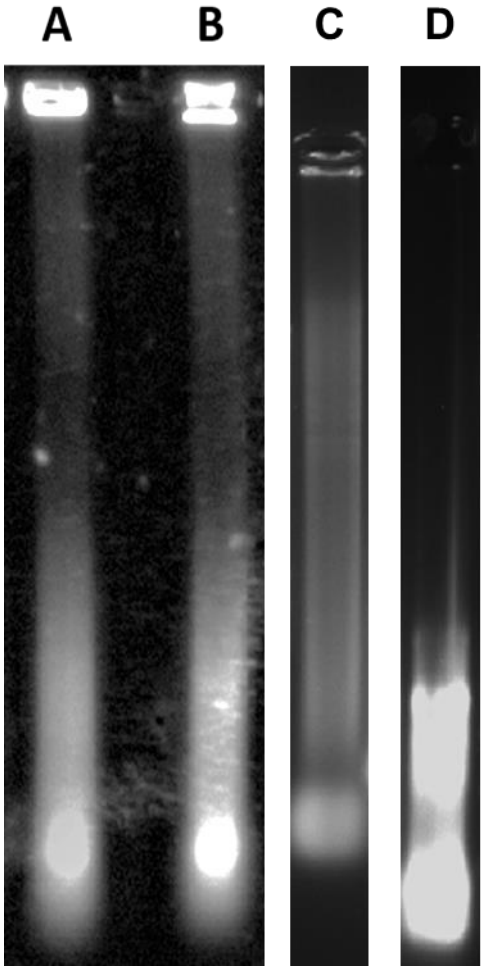

Figure 4-2 Comparison of nucleic acid extracted (A) and sonicated (B) *P. pastoris* gDNA integrity. (C) is an example of un-sheared DNA and (D) is an example of sheared DNA, for comparison.

4.2.2 Comparison between boiling and sonication sample preparation methods

Rapid sample preparation for PCR study involves the lysis of all cells within a sample, in order to release the nucleic acid analyte. Traditional sample preparations achieve this by isolating then nucleic acid through chemical or physical means; however their impact on assay throughput is great. We initially sought to compare two methods of rapid sample preparation, a 10 minute boiling protocol and a 5 minute total sonication protocol. To achieve a comparison we divided a sample in half and treated one by sonication and the other by boiling. We then compared each method by comparing the Cq values obtained from qPCR analysis across three dilutions. Should the Cq values be equivalent, it can be determined that the impact on efficiency through inhibition is negligible. Significantly lower Cq values are indicative of efficiency being lowered through inhibition. This was carried out over three technical replicates and the average Cq gained was recorded.

We found, as shown in Figure 4-3, that the two methods of rapid sample preparation are approximately equivalent, as demonstrated by the average Cq values being approximately equivalent. As we are looking to achieve the highest possible assay throughput, we chose the sonication method due to it being the most rapid method.

4.2.3 Influence of disrupted *P. pastoris* cells on e-pPCR sensitivity

We sought to quantify the degree to which the presence of *P. pastoris* cellular material influences the sensitivity of e-pPCR. To achieve this we used primers generated using the design strategy outlined (Figure 3-2) specific for a transketolase genomic target. In order to disrupt cells and liberate host gDNA whilst minimising gDNA shearing or degradation, we used a brief and mild sonication procedure. Agarose gel electrophoresis showed that no discernible shearing of gDNA had occurred post sonication compared to pre-sonication samples (Figure 4-2). As an example, C shows sheared DNA, where the analyte has moved to the bottom of the gel due to the reduced size of the DNA fragments. To quantify the effect of total disrupted cellular material, e-pPCR was performed using either purified gDNA or disrupted cells as template. We defined the limit of detection (LOD) as that tenfold dilution of template material which resulted in

no detectable amplicon band after n=3 experimental repeats. To allow direct comparison of the effects of cultivation method, DNA template mass and cell numbers were matched between shake flask and bioreactor samples (Figure 4-4).

The presence of shake flask cellular material reduced the total production of amplicon band, summed from all reactions, by 48% (from 101 to 52 arbitrary units) compared to the purified, gDNA-only template (Figure 4-4 A). By contrast, cellular material from bioreactor cultivation (Figure 4-4 B) reduced template amplification by only 18% (from 189 to 154 arbitrary units) and a greater overall level of template amplification was observed when compared to shake-flask derived material.

The observed LOD was 50pg for purified gDNA for both shake flask (Figure 4-4 A) and bioreactor (Figure 4-4 B) samples. The presence of cellular material reduced e-pPCR sensitivity sixfold to 300pg gDNA for both shake flask and bioreactor samples.

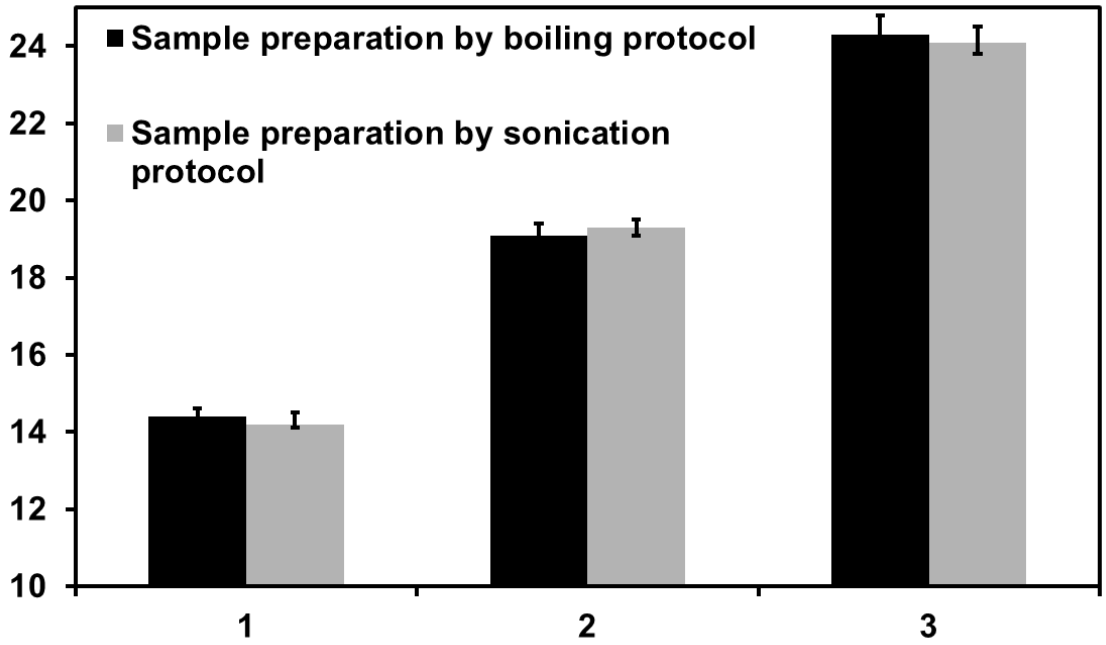

Figure 4-3 Comparison of boiling and sonication sample preparation protocols in terms of impact on Cq value. N=3 experiments.

4.2.4 Influence of disrupted *P. pastoris* cells on qPCR amplification efficiency

A plethora of approaches to qPCR data analysis are used across many fields of science and engineering. However, a practice common to many methods is to calculate the efficiency of amplicon production as an indicator of accuracy (Pfaffl, 2001). Typically a threshold of 100±10% efficiency, at $R^2 > 0.99$ (Gil and Coetzer, 2004), must be satisfied for a data point to be considered accurate.

We compared the efficiency of amplification when purified genomic DNA (gDNA) and disrupted cellular material are used as template (Figure 4-5), as an indicator of the degree to which sample preparation is necessary in qPCR-based procedures. We defined limit of quantitation (LOQ) as that template dilution for which the coefficient of determination in an efficiency of 100±10% falls below $R^2 = 0.99$. For purified gDNA analysis the samples indicated in Figure 4-1 underwent total gDNA purification followed by resuspension in dH_2O to their original sample volume. For disrupted cellular material, the samples indicated in Figure 4-1 underwent centrifugation and suspension to the original volume with dH_2O before sonication. All samples were then tenfold diluted as indicated in Figure 4-5 and used as template for qPCR.

For shake flask samples (Figure 4-5), pure gDNA enabled 100±10% amplification efficiency, with $R^2 > 0.99$, over 6 template dilutions. The presence of cellular material reduces this to 4 dilutions, decreasing the LOQ by two orders of magnitude. Surprisingly, the equivalent experiment with bioreactor samples (Figure 4-5), revealed amplification of both pure gDNA and disrupted cell material as template at 100±10% efficiency over 6 or more dilutions. Tenfold dilution of the initial OD_{600}=800 bioreactor sample, down to OD_{600}=80, is required, after which the presence of yeast cell material has little effect on assay performance.

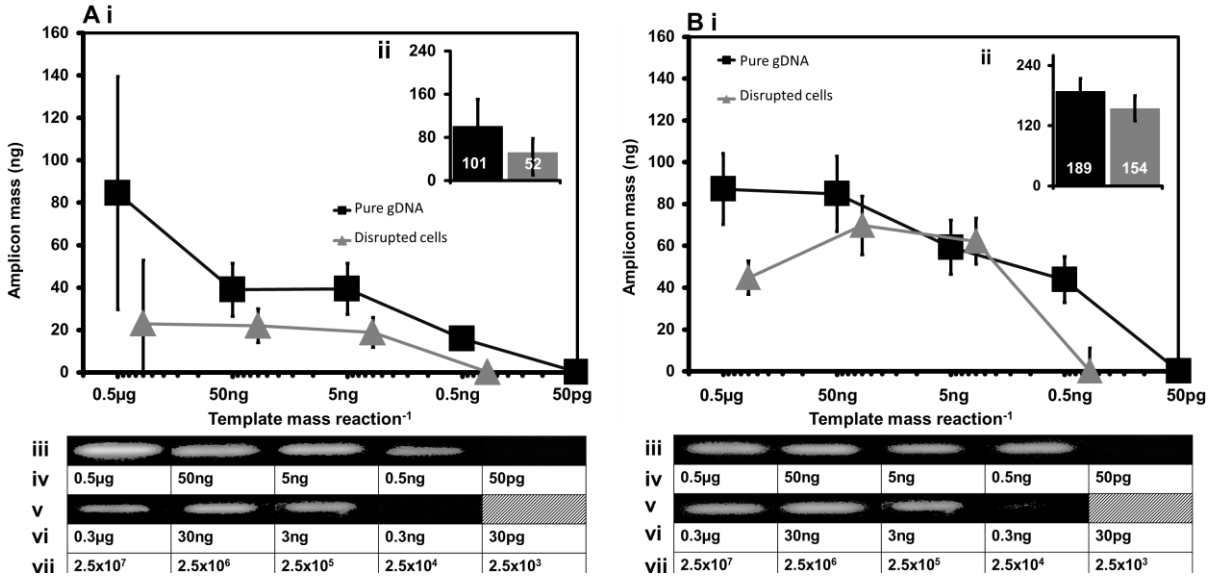

Figure 4-4 Influence of disrupted cells on e-pPCR sensitivity. PCR and subsequent gel electrophoresis was performed using disrupted cells (gray symbols and lines) from shake flask (A) and bioreactor (B) cultivation as template material. Duplicate cell samples in which total gDNA had been purified away from cellular material was also used as template (black symbols and lines). For panels A and B, graph i) indicates densitometry measurements of resultant 104bp amplicon band. Inlaid graph ii) plots the area under each curve in graph i) with the total value indicated in white text within the bar. For reactions with pure DNA as template, gel photos (row iii) and estimated template mass (row iv) are shown. For reactions in which disrupted cells were used as template, gel photos (row v), estimated template mass (row vi) and estimated cell numbers present pre-sonication (row vii) are shown.

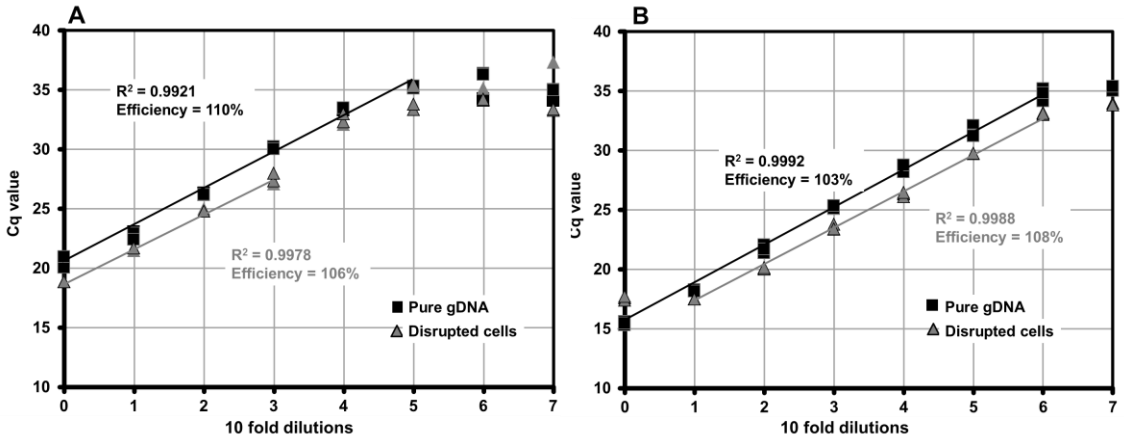

Figure 4-5 Influence of disrupted cells on qPCR efficiency. Quantitative PCR was performed using either disrupted cells (gray symbols and lines) or purified gDNA (black symbols and lines), from shake flask (A) and bioreactor (B) cultivation, as template. Data sets are representative of n=3 analytical repeats.

4.2.5 Influence of disrupted P. pastoris cells on 'Standard Curve' and 'LRE' qPCR

Rutledge et al. demonstrated a novel approach to qPCR using a method of linear regression of efficiency (LRE) for analysis of fluorescence data. LRE-qPCR also features what Rutledge et al. (2008) refer to as a universal standard reaction, the optical calibration factor (OCF), consisting of a defined 'Cal1' primer pair plus lambda phage DNA as template. Rutledge et al.(Rutledge and Stewart, 2010) report that the Cal1 reaction exhibits near-ideal behaviour of 100% amplification efficiency, enabling highly accurate correlation of amplification performance with the appearance of fluorescence. This approach does not require any relatedness between the Cal1 standard reaction and the experimental reaction. The authors tested LRE-qPCR with the Cal1 OCF over a four-month period with no significant change in OCF performance (Rutledge and Stewart, 2010). We suggest these inherent properties of the Cal1 OCF standard make it ideal for use as a synthetic biology standard for application of qPCR to *P. pastoris* samples. To validate LRE-qPCR with the Cal1 OCF as a synthetic biology standard we compared its accuracy with that of a conventional Standard Curve (SC) method of qPCR.

Shokere et al., (2009) showed that spectrophotometry and the standard curve qPCR (SC qPCR) provide comparable DNA concentration measurements when used with purified DNA samples. Counter to our expectations, we also observed that spectrophotometry could be used to measure DNA concentration even when used with crude suspensions of disrupted cells from samples of up to OD_{600}=80, despite the presence of many components likely to distort the absorbance spectra (see Table 11) and typical absorbance profiles shown in Figure 4-6). As such we used spectrophotometry as a mechanistically distinct comparator method to assess the performance of both LRE-qPCR and SC qPCR.

Table 11 Spectrophotometric DNA measurements and indication of DNA loss in purification step.

Source	**Dilution**	**OD600**	**DNA (ng/µL) in disrupted cell solution**	**DNA (ng/µL) DNA purified from cells**	**DNA Loss (%)**
Shake Flask	0	50	776.1	319	58.9
	1	5	79.4	31	60.96
	2	0.5	7.7	3.2	59.09
				Mean	**59.65%**
Bioreactor	1	80	272.9	174.1	36.2
	2	8	30.9	17.4	43.69
	3	0.8	3.1	1.7	45.16
				Mean	**41.68%**

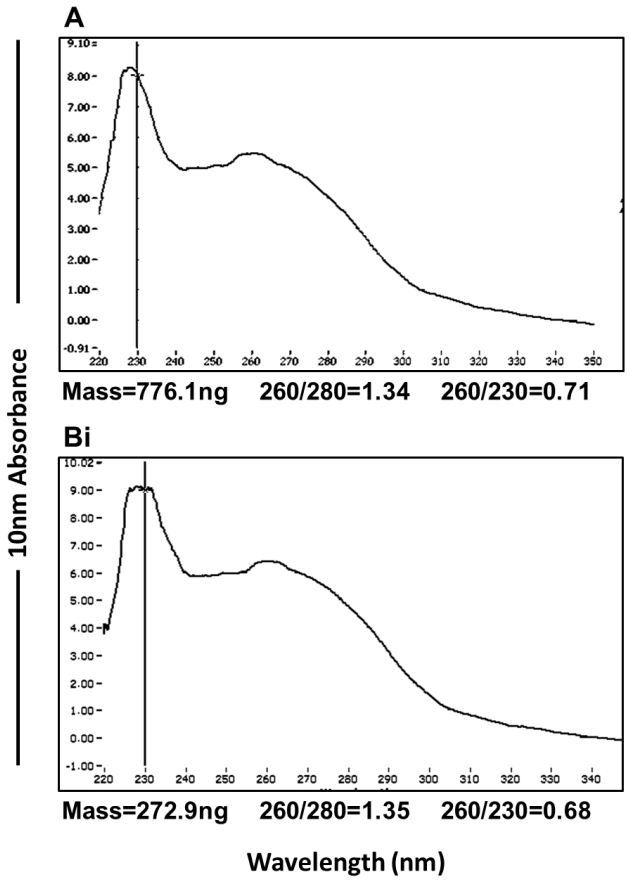

Figure 4-6 Spectrophotometric measurements using disrupted cell suspensions. DNA absorbance profiles of disrupted cells derived from an OD600=50 shake flask sample (A) and an OD600=80 bioreactor sample (B) using a Nanodrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA).

Spectrophotometrically determined DNA mass in disrupted cellular material and pure DNA samples (Figure 4-6) was used to predict genome copy number based on a *P. pastoris* genome size of 9.43 Mb (De Schutter et al., 2009). As the transketolase target locus is known to be present as a single copy within the *P. pastoris* genome, genome copy number is assumed to be equivalent to target sequence copy number. As such we converted three spectrophotometer measurements to target sequence copy numbers). We plotted these three spectrophotometrically derived target sequence copy numbers (grey circles in Figure 4-8) as a function of template dilution and linearly extrapolated the trend (dashed lines) for both shake flask (Figure 4-8 A) and bioreactor material (Figure 4-8 B).

For the conventional SC qPCR method the 'standard curve' used for calibration is generated by linear regression of Cq values obtained with log dilutions of purified samples of template DNA. Figure 4-7 shows the standard curves generated for both shake flask and bioreactor quantification and Figure 4-8 compares the SC and LRE-qPCR methods of quantification.. The template DNA used for this standard curve must be purified, of known concentration and also the same sequence as the DNA expected to be present in the experimental samples at unknown concentration. As such, purified gDNA samples, where concentration has been measured by spectrophotometry, represents the standard curve for the SC qPCR method. For the LRE-qPCR method the proposed Cal1 universal OCF standard was used to calibrate the data and linear regression is applied directly to the fluorescence curve for every cycle of the reaction (Figure 4-8 D).

For shake flask material (Figure 4-8 A), the SC method was able to quantify target DNA in undiluted (OD_{600}=50) cellular material and over two further tenfold dilutions after which copy number values diverged from the spectrophotometric data. At the fifth dilution of the standard curve data, the individual data points diverge around the quantity predicted from the spectrophotometry data. In order to confirm statistical significance a t-test was performed, which compared the individual data points to the spectrophotometry data. This resulted in a p value of 0.17, indicating no statistical significance and a probable divergence of quantification by SC analysis from that

extrapolated by spectrophotometry. The LRE-qPCR method produced results in agreement with spectrophotometric data for undiluted material but also over three further tenfold dilutions and with less divergence of individual data points.

For bioreactor material (Figure 4-8 B), both LRE-qPCR and SC qPCR methods are in close agreement with spectrophotometrically derived target DNA copy number (Table 11) for material that has undergone one tenfold dilution (therefore OD_{600}=80) and five further tenfold dilutions.

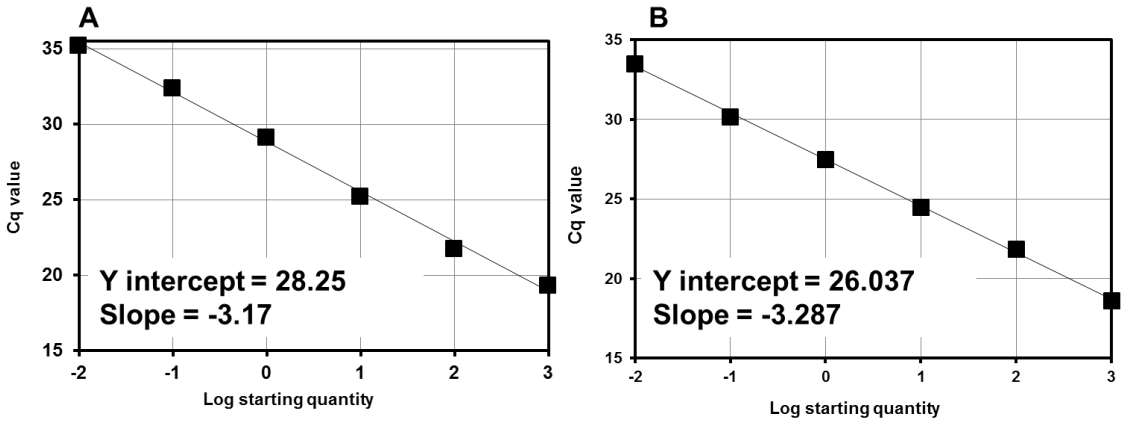

Figure 4-7 Standard curves used in quantification of Cq values generated from sonicated *P. pastoris* process streams. (A) shows the standard curve generated from purified shake flask material. (B) shows the standard curve generated from purified bioreactor material.

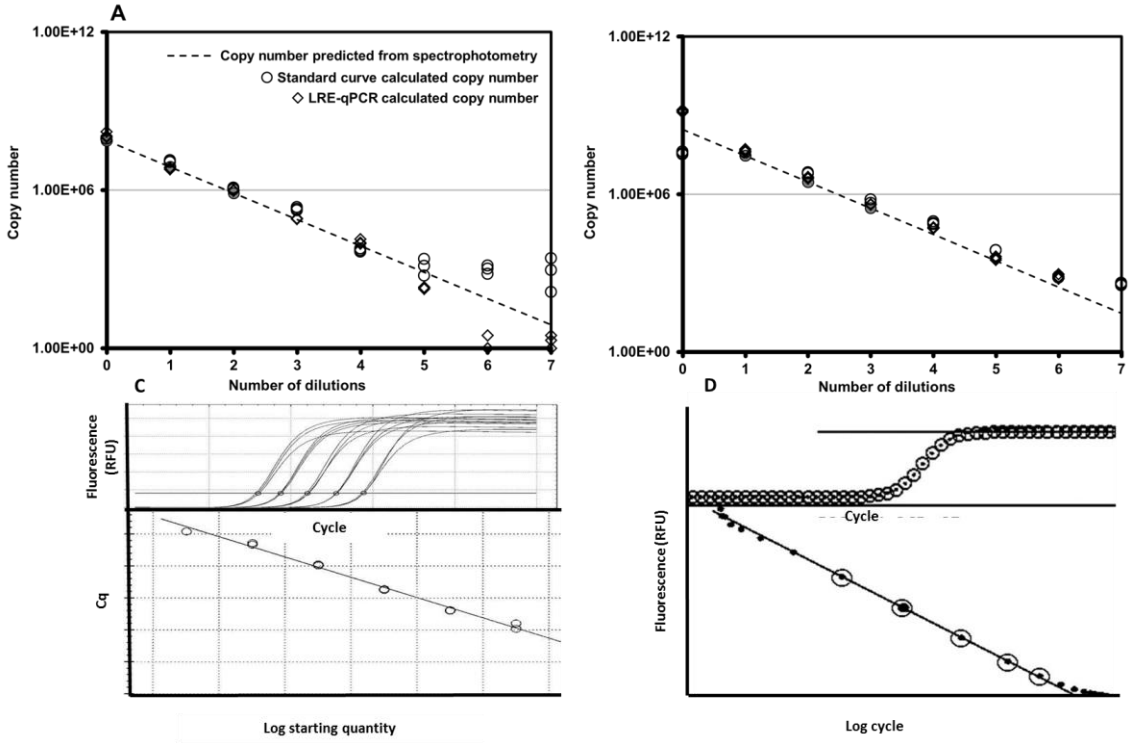

Figure 4-8 Comparison of SC qPCR and LRE-qPCR using disrupted cells as template. Both graph A (shake flask) and graph B (bioreactor) incorporate the following features. Grey data points indicate spectrophotometric data. The absorbance profiles for the disrupted cell sample derived from OD600=50 shake flask culture (A) and OD600=80 bioreactor culture (B) are shown in Figure 4. Dashed lines linearly extrapolate the spectrophotometric data points to predict copy number at lower dilutions. Open circles indicate copy number of target present in disrupted cells as determined by SC qPCR. Open rhomboids indicate copy number of target present in disrupted cells as determined by LRE-qPCR. Graph C shows the standard curve used for SC qPCR quantification, formed by plotting the Cq values (open circles) of multiple reactions as a function of the log of their serial dilution. Raw fluorescence data and a horizontal 'crossing threshold' line is shown in the inlaid graph above the Cq data. In graph D fluorescence values (circles) of the Cal1 OCF reaction are plotted against their cycle number (upper panel) and the log of their cycle number (lower panel). Data sets are representative of n=3 analytical repeats.

4.2.6 Statistical comparison of 'Standard Curve' and 'LRE' qPCR methods

We used an XY plot (Burd 10) to compare SC qPCR and LRE-qPCR for quantifying levels of target DNA in disrupted cellular material. A slope of 1.00 indicates zero bias between methods. XY plot showed that, for shake flask material, LRE-qPCR (

Figure 4-9 A) showed marginal proportional bias (slope of 0.914) of SC qPCR. The Y intercept for this comparison did show large deviation from zero (0.502), which would suggest real systemic bias of SC qPCR. However, Bland-Altman bias plots (Bland and Altman, 1986) indicated that LRE-qPCR and SC qPCR methods are equivalent for analysis of shake flask derived samples (

Figure 4-9B), as the mean bias spanned the zero difference level and generally fell within confidence limits (Burd, 2010).

For bioreactor-derived samples, an XY plot of SC qPCR and LRE-qPCR data revealed a slope of 1.1758, indicative of only modest proportional bias (

Figure 4-9C). Although again the Y intercept for this comparison did show large deviation from zero (-0.8513), a Bland-Altman bias plots was consistent with the LRE-qPCR and SC qPCR methods being equivalent as the mean bias level spanned zero difference (

Figure 4-9D).

4.2.7 LRE-qPCR with Cal1 OCF as a synthetic biology standard for qPCR in *P. pastoris*

Unlike SC qPCR, LRE-qPCR does not require a standard curve consisting of indentical primers and target as experimental samples of unknown target DNA concentration. This allowed us to measure the profile of LRE-qPCR accuracy for both purified gDNA and disrupted cellular material (Figure 4-10). LRE-qPCR data in Figure 4-10 was calibrated using the Cal1 OCF primers and target. As previously, we used spectrophotometric data

to predict genome copy number present in a given sample. Shake flask data in Figure 4-10A shows the effect of the presence of cell material from shake flask on the ability of LRE-qPCR to match the spectrophotometric prediction, with agreement for undiluted material and over three subsequent tenfold dilutions. For HCD bioreactor material (Figure 4-10B) LRE-qPCR also matched spectrophotometric data across all samples except for the most concentrated sample and most dilute samples, neither of which yielded amplicon.

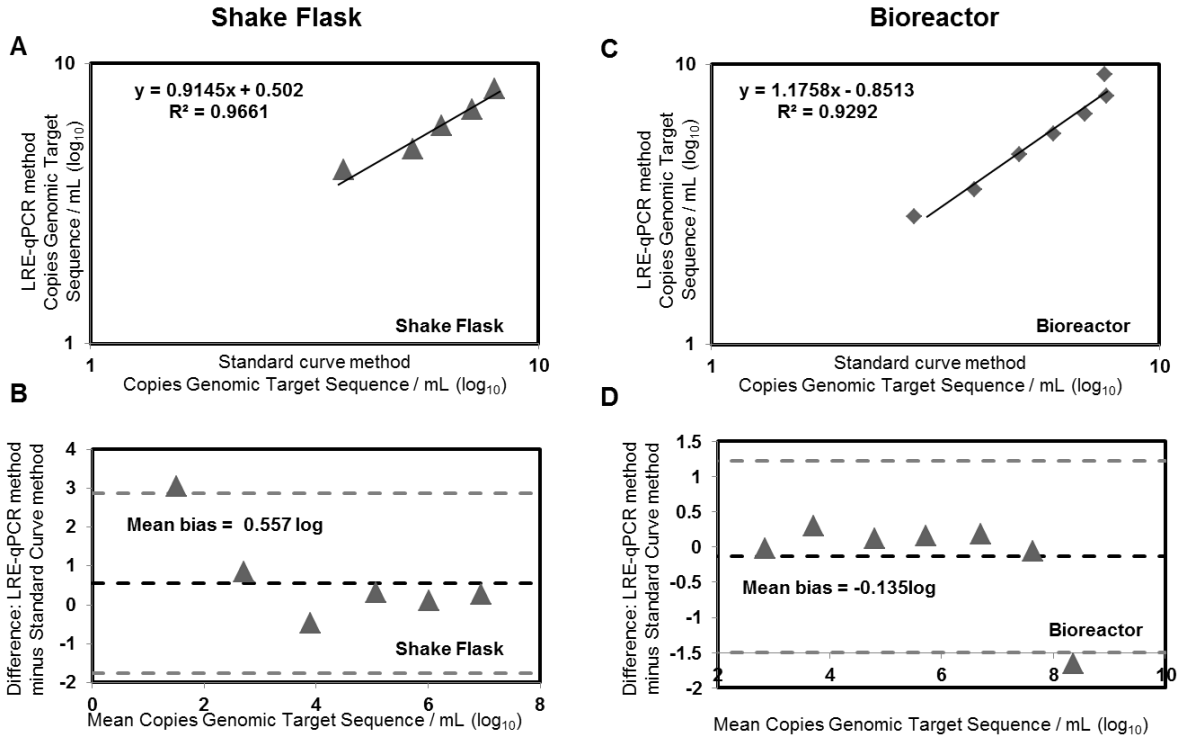

Figure 4-9 Statistical comparison of SC qPCR and LRE-qPCR. XY plots (graphs A and C) were derived from copy number estimations made using the indicated method, using data plotted in graphs A and B of Figure 5. Bland-Altman bias plots (graphs B and D) were derived from XY plots. Statistical procedures were performed as described by Burd (2010).

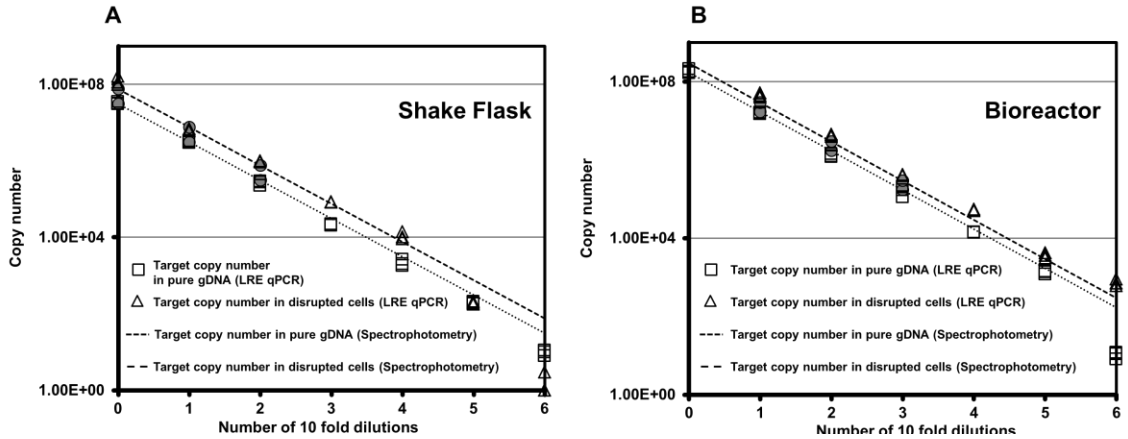

Figure 4-10 LRE-qPCR performance using pure gDNA and disrupted cells as template. LRE-qPCR was performed on samples derived from shake flask (A) and bioreactor (B) cultures. Grey data points indicate spectrophotometric data and dashed lines extrapolate these data to predict copy number at lower dilutions. LRE-qPCR data is plotted with square symbols for purified gDNA and triangles for disrupted cells. Data sets are representative of n=3 analytical repeats.

4.3 Discussion

Sample preparation is normally performed to remove inhibitors that might impact the accuracy and sensitivity of PCR-based assays (Dineva et al., 2007). Ionic detergents, phenol and metal salts that may be present in growth media can all inhibit PCR. Kits and reagents used to extract, purify and preserve DNA can also influence PCR and bring the risk of introducing error through loss of DNA (Miller et al., 1999), introduction of inhibitory biological material from disrupted cells and co-purification of chemical inhibitors (Schrader et al., 2012). Some DNA isolation kits have also been shown to produce false-positive results due to the presence of contaminant DNA in the kit, the level and source of which varies between manufacturer and batch (Queipo-Ortuño et al., 2007).

End-point PCR remains a valuable tool for detecting contaminant organisms during scale up or storage of biological material. One factor that delimits e-pPCR application is the time it takes to perform sample preparation. We sought to measure the extent to which sample preparation is necessary for garnering a reliable yes / no binary datum using e-pPCR (Figure 4-4). Cell suspensions from shake flask or bioreactor cultivation were sonicated for 30s as part of a procedure that took a total of 5 minutes to perform before being used in e-pPCR.

As expected, significant variation in the level of amplicon production was observed between experimental repeats - underlining the fact that e-pPCR is best suited to detection and not quantitation. The total the mass of amplicon produced was greater for bioreactor material than for shake flask material even though the templates had been matched in terms of number of cells and gDNA mass (see Materials and Methods). Despite the difference in amplicon production, both shake flask and bioreactor material reduced e-pPCR sensitivity six fold (Figure 4-4). These observations suggest that *P. pastoris* sample preparation for binary e-pPCR assays is necessary to avoid a significant reduction in LOD.

In contrast to e-pPCR, qPCR is used widely to accurately quantify relative or absolute abundance of DNA. Polymerisation and analysis occur in parallel in most qPCR platforms, unlike e-pPCR in which gel electrophoresis and gel analysis are performed serially after PCR. Sample preparation therefore represents a greater proportion of total assay time (Figure 4-11) for qPCR. As such we characterised the influence of shake flask and bioreactor samples on qPCR efficiency again using sonication to represent a non-processed sample in which DNA and all other cellular components remain present throughout polymerisation.

LRE-qPCR is calibrated against the Cal1 reaction so we could directly test its performance using purified gDNA samples and disrupted cells as template. This allowed us to map the influence of shake flask and bioreactor material on LRE-qPCR analysis. Figure 4-5A data shows that shake flask material does compromise the LOQ for LRE-qPCR but Figure 4-5B indicates that bioreactor material has minimal effect on LRE-qPCR, except for the most concentrated (originating from OD_{600}=800) and the most dilute samples.

Commercial sample preparation kits, including membrane and bead-based systems, often involve protocols that comprise 20 steps or more and can take over one hour to perform. Our findings suggest that a simple and rapid (approximately 5 minute) sonication procedure is sufficient to render HCD bioprocess samples amenable to LRE-qPCR analysis. Current (http://www.roche.com/products/product-details.htm?type=product&id=64) and prototypical (www.xxpresspcr.com) ultra-rapid PCR devices offer the potential to reduce reaction and analysis time to less than 20 minutes. A 30 minute or less total procedure duration would pose the real possibility of LRE-qPCR being used as an at-line bioprocess monitoring tool and as such offer new analytical power for process development. To illustrate this we logged the length of time taken to complete several of the procedures undertaken during this work and used these data to project likely future assay durations in Figure 4-11.

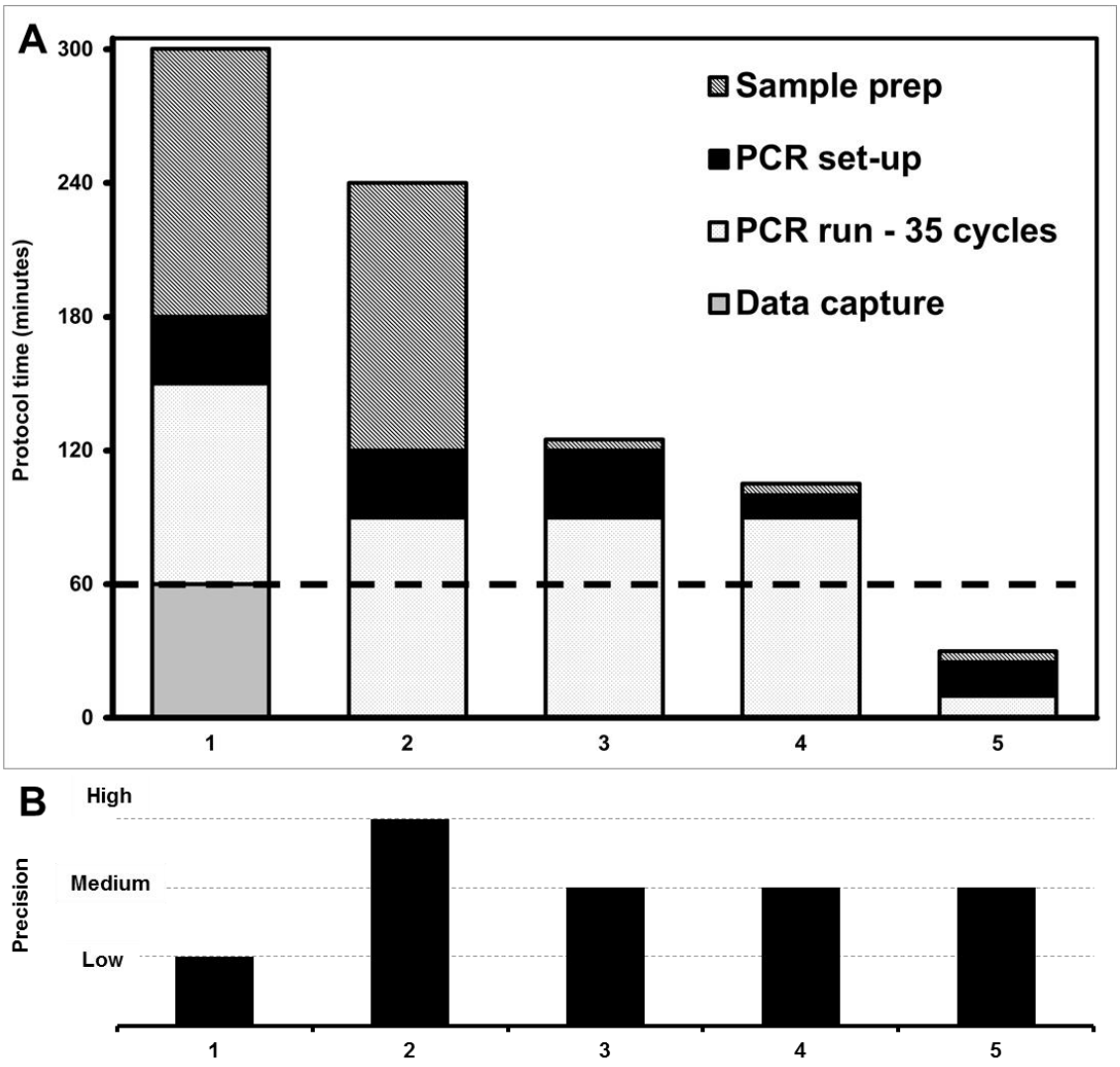

Figure 4-11 Predicted and known time profiles of different PCR methods. Data collected from this work (see Materials and Methods) and provided by manufacturers is used to compare typical durations of current and predicted PCR protocols. Assays of 60 minutes duration or less (dotted line) are compatible with at-line bioprocess monitoring. (A) X axis numbers indicate observations for the following procedures: 1) e-pPCR with agarose gel data capture, 2) Standard Curve qPCR using a BioRad CFX qPCR device with parallel fluorescent measurement and analysis. X axis numbers indicate predictions for the following procedures: 3) Standard Curve qPCR using a rapid sample preparation procedure such as the sonication step in this work, 4) LRE-qPCR with rapid sample preparation and 5) LRE-qPCR with rapid samples preparation and run on an ultra-rapid device such as the Xxpress (BJS Biotechnology) or the Roche LightCycler. (B) is an estimate of precision based on bias of data, assay stages and variability of data gained

Why does the presence of shake flask material affect both e-pPCR (Figure 4-4) and qPCR (Figure 4-4, Figure 4-8, Figure 4-10) more than high cell density bioreactor material? This may be due to the physiological status of cells 66 hours post-induction and at the end of idiophasic cultivation, compared to their pre-induced state during seed train, shake flask growth (Figure 4-1). To maximise product yield many microbial fermentations typically proceed for durations that can cause significant physiological impact on cells (Sandén et al., 2003). Levels of dyshomeostasis and misfolded protein accumulation in late idiophase cells may have the net effect of making cellular material more readily unbind DNA than is the case for healthier cells. The observation that phenol/chloroform DNA purification from shake flask material is loses more template material than purification from high cell density bioreactor material (Table 11, and plotted in Figure 4-10), is consistent with this hypothesis.

Most current efforts in qPCR standards deal with experimental setup (Bustin et al., 2009), food (Malorny et al., 2003) or water safety testing, with several standards (Table 12) agreed by the International Standards Organisation (ISO). We suggest that, due to the advantages of having Cal1 OCF standard, LRE-qPCR need only match the accuracy of conventional SC qPCR in order to be a credible standard for bioindustry and the synthetic biology community. Figure 4-8 shows that, for both shake flask and bioreactor material, LRE-qPCR matched SC qPCR in ability to confirm copy number predictions made by extrapolation of spectrophotometric data. Head-to-head comparison also showed the methods to be equivalent (Figure 4-9).

Table 12 Comparison of calibration and standardisation for LRE and SC qPCR. Rutledge and Stewart (2010) observed negligible variation in the performance of the CAL1 OCF over 8 runs across a 4-month period. We interpret this observation as a strong indicator that calibration runs may only be necessary for LRE-qPCR is infrequently as once every 4 months.

	Calibration Method	**Recommended Frequency of calibration**	**Proposed standards**
Standard Curve	Parallel reactions of samples containing known DNA mass (Higuchi et al., 1993).	Every assay (Higuchi et al., 1993).	ISO 22119:2011 (De Schutter et al. 2009)(De Schutter et al., 2009). ISO 22119:2011 (De Schutter et al. 2009)(De Schutter et al., 2009). ISO/TS 13136:2012. ISO/TS 12869:2012. ISO/TS 21569-2:2012. Standard proposed by Malorny et al. 2003.
LRE	OCF generated from the CAL1 lambda DNA calibration reaction.	Every 4 months (Rutledge and Stewart 2010).	This report.

Assay duration is a key delimiting factor at present in the application of qPCR for monitoring industrial processes. In future, assay duration is also likely to be a key factor when monitoring the status of performance-critical loci within synthetic yeast genomes or gene networks (Guo et al., 2015). A significant advantage LRE-qPCR brings is the reduced need for calibration runs, with the Cal1 reaction recommended to be performed only every 3-4 months as opposed to Standard Curve calibration which is ideally performed for every experiment.

We have shown that sample preparation is a critical requirement e-pPCR analysis of *P. pastoris* material from both shake flasks and bioreactors. By contrast, for qPCR analysis a simple and rapid sample preparation stage of media removal and suspension in hH2O followed by sonication, with no attempt at DNA purification, is sufficient to gather accurate qPCR data from HCD bioreactor material. However, it is likely that a degree of DNA purification is necessary for accurate qPCR analysis of shake flask-derived *P. pastoris* material.

LRE-qPCR has inherent advantages in terms of standardisation and the frequency of required calibration reactions. LRE-qPCR matches conventional Standard Curve qPCR with respect to absolute quantification of target DNA, even in the presence of OD_{600}=80 material. We predict a combination of rapid sample preparation, adoption of the LRE-qPCR Cal1 OCF standard and devices capable of ultra-rapid PCR will enable expansion of qPCR to at-line monitoring of yeasts controlled by synthetic genomes at scale. We invite the synthetic biology and biotechnology communities to test further the Cal1/OCF standard and LRE-qPCR method for absolute quantitation of genomic sequences in *P. pastoris* and to assess the procedure as a standard for use with other organisms.

4.4 Conclusions

- Sample preparation is a requirement e-pPCR analysis of P. pastoris material from both shake flasks and bioreactors.
- A simple and rapid sample preparation stage is sufficient to gather accurate qPCR data from HCD bioreactor material, although it is likely that a degree of DNA purification is necessary for accurate qPCR analysis of shake flask-derived P. pastoris material.
- LRE-qPCR shows comparable accuracy to SC qPCR analytical methods, and we propose its CAL1 standard as a Synthetic Biology complaint standard for qPCR

- With the methodological changes developed here, we demonstrate that this assay has the potential for at-line qPCR analysis if applied to ultra-rapid thermocycling platforms

Chapter 5

5 Influence of high cell density Chinese Hamster Ovary (CHO) cells on performance of PCR as a synthetic biology tool for bioprocess monitoring and contaminant detection

5.1 Introduction

As seen by developments in genetic manipulation and industrial utilisation of yeasts, the design and construction of synthetic eukaryotic genomes has made rapid progress in recent years (Dymond et al., 2011), alongside conventional recombinant DNA approaches to construction of human artificial chromosomes (Kononenko et al., 2015). It is logical to expect these two research trajectories will converge with the design, construction and implementation of synthetic genomes to control mammalian cells. This holds the promise of powerful control of those mammalian cell characteristics that currently limit their performance in industrial settings. One challenge raised by this vision is the need for standardised assays to quantify the presence or loss of operationally critical genetic elements within a synthetic genome. Standardisation of data capture metrics is a defining feature of both synthetic biology and industrial bioprocessing and is critical to reproducible manipulation of host chassis (Kitney and Freemont, 2012). Data captured during industrial scale cell cultivation is also essential to achieve process understanding, which in turn is necessary for optimisation of product yield and quality (Clementschitsch and Bayer, 2006).

As mammalian cells have the capacity to produce recombinant proteins with human-like glycosylation and post translational modification, they are essential for the production of therapeutic monoclonal antibodies (mAbs), (Walsh and Jefferis, 2006). The Chinese Hamster Ovary (CHO) cell production platform is one of the most widely researched and exploited chassis for the production of biologics such as mAbs (Kim et al., 2011). Mycoplasmal infections of CHO and other mammalian cells types can distort cell phenotype, compromise host genome integrity (Lincoln and Gabridge, 1998) and

confound efficacy of cell-based therapies (FDA - Food and Drug Administration and Center for Biologics Evaluation and Research, 2010). As such mycoplasmal infection is a major risk factor that potentially jeopardises clinical translation of many of the exciting advances made by synthetic biologists in areas such as T cell therapy (Nikfarjam and Farzaneh, 2012). PCR and its variants has become the gold standard for detecting infection and profiling genome integrity of CHO cell lines.

Whilst there are a number of PCR-based assays utilised in mammalian culture systems, there is little-to-no level of standardisation between methodology and nucleotide sequences. Furthermore, the same sample preparation procedures are used that typically extend assay throughput time, increase labour and can introduce error (Skulj et al., 2008). This presents a particular limitation for industrialists seeking to gain insight from approaches such as process analytical technology (PAT) which ideally require real-time or at-line analysis (Kaiser et al., 2008).

With these challenges in mind we suggest it is prudent to define some of the factors involved in developing rapid, industrially robust and standardised PCR-based assays for monitoring of both CHO cell genomic loci and contaminant organisms known to be a risk factor in CHO cultivation. To do this we quantitate the impact of CHO cells, from shake flask and bioreactor cultivation, on the performance of PCR. We test end point PCR (e-pPCR), a conventional method of standard curve-based (Pfaffl et al., 2002) quantitative PCR (SC qPCR) and the recently-developed linear regression of efficiency qPCR method (LRE-qPCR), which features a calibration reaction (Rutledge and Stewart, 2010) reported to have ideal amplification properties that enable its use as universal standard. We believe the resulting data will reveal the extent to which sample preparation is in fact required for PCR, if at all, enabling future efforts by ourselves and others to developme a rapid, robust and standardised PCR assay. We also continue to examine our findings in light of possible application of the LRE-qPCR calibration reaction as a synthetic biology standard applicable to CHO cell fermentation.

5.2 Results

5.2.1 Cultivation of CHO cells by shake flask and bioreactor

We used a commercial CHO cell line that constitutively expresses and secretes a therapeutic monoclonal antibody. Cells were grown in CD-CHO media (Invitrogen, CA, USA) in 1L Erlenmeyer shake flasks to a cell density of $2x10^6$ cells / mL. At this point a sample was taken for further PCR experiments (Figure 5-1 A). This stage represents a critical point of industrial scale cultivation where the seed train is used to inoculate the larger scale growth vessel. The shake flask culture was added to CD-CHO media in a rocked bag bioreactor to an initial concentration of $2.5x10^5$ cells/mL. Cells then grew to achieve $1x10^{11}$ cells / mL with ~99% viability (Figure 5-1 B) and a further sample was removed for PCR. This cell concentration is typical of the point before harvest and product recovery when contamination from infectious agents presents the most significant economic burden.

5.2.2 Preparation and analysis of material containing template DNA for PCR

To evaluate the effect of cellular material on PCR assay performance we disrupted cells using a gentle sonication procedure configured to ensure gDNA remained largely intact and was not denatured to any significant degree. This was confirmed by agarose gel electrophoresis of cell suspensions before and after sonication (Figure 5-2). Disrupted cell samples were compared to samples in which total nucleic acids had been isolated using standard phenol-chloroform extraction. Spectrophotometry was used to determine DNA mass to enable genome copy number estimation by a method that is mechanistically unrelated to PCR. Three spectrophotometric measurements were taken over three tenfold serial dilutions and this was used to predict DNA mass over further dilution. Densitometric analysis of gel images was also used to estimate total DNA concentration present in a given sample of disrupted cells.

5.2.3 CHO cellular material reduces e-pPCR sensitivity tenfold for detection of a genomic locus

As well as contaminant detection, end-point PCR (e-pPCR) is widely used to confirm the identity of a host cell by confirming the presence of a single genomic locus (Parodi et al. 2002). We quantified the impact of disrupted cells, from shake flask and 3L scale bioreactor cultivation, on the limit of detection (LOD) for e-pPCR used to confirm the presence of a sequence within the single genomic copy GAPDH gene. Template material consisted of either total nucleic acid purified from a cell suspension sample, or disrupted cells. LOD was taken to be the tenfold dilution of template material for which no amplicon band could be detected after n=3 experimental repeats. The presence of disrupted cells lowered e-pPCR LOD by one order of magnitude for samples from both shake flask (Figure 5-3 A) and bioreactor (Figure 5-3 B). Marginally lower overall amplicon production was observed in samples taken from bioreactor cultivation (Figure 5-3 Bii) compared to shake flask (Figure 5-3 Aii).

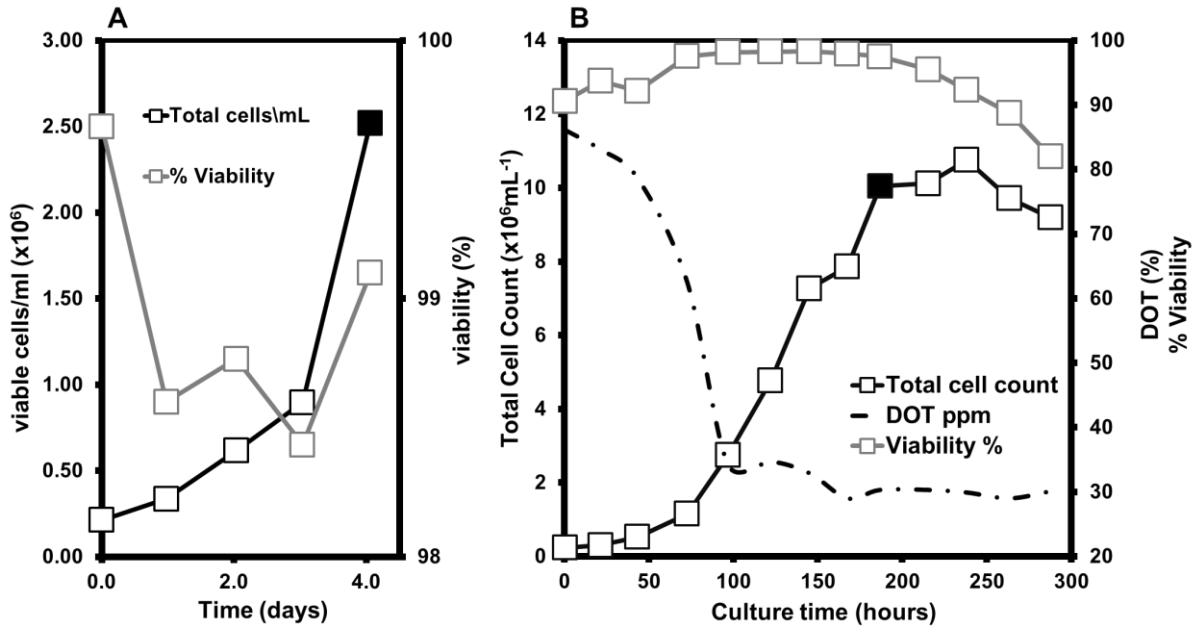

Figure 5-1 Cultivation of CHO cells. Growth profile of CHO cells in T-175 flasks (A) and in an Applikon 3L rocked bag bioreactor fermentation (B). Samples for PCR experiments were taken at the time point indicated (closed symbol).

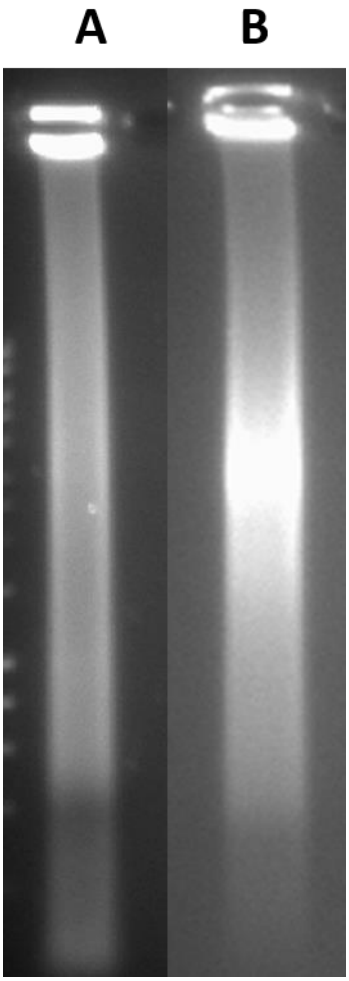

Figure 5-2 Comparison of nucleic acid extracted (A) and sonicated (B) CHO gDNA integrity

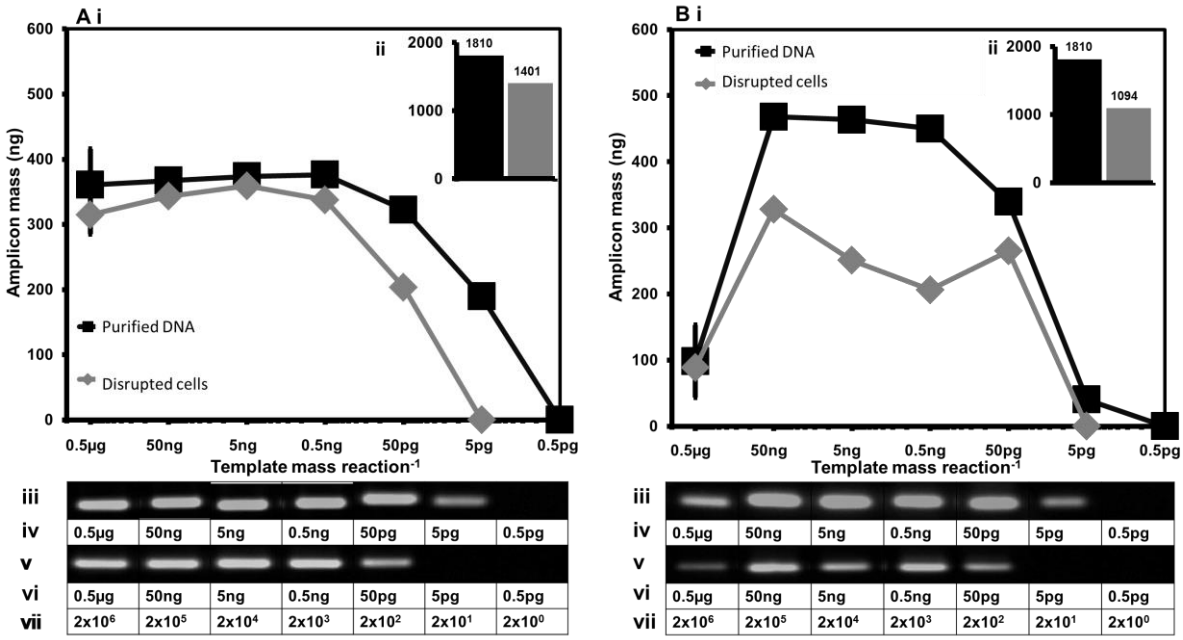

Figure 5-3 Influence of disrupted CHO cells on e-pPCR detection of a genomic target sequence. Disrupted cells and purified DNA from samples taken from shake flask (A) and bioreactor (B) cultivation were used as template material for e-pPCR. For both cultivation methods the following data are depicted. The mass of amplicon produced in a reaction is plotted as a function of sample dilution (i). Inlaid graphs (ii) plot the area under each curve as a bar chart. Agarose gel images show the amplicons generated from the purified DNA (iii) and disrupted cells (v). Template DNA mass in purified DNA samples (iv) and disrupted cell samples (vi) was measured by spectrophotometry. The numbers of cells sonicated to generate the disrupted cell samples are indicated in the row labelled (viii). Error bars represent standard deviation over n=3 experimental repeats.

5.2.4 Efficiency of genomic target amplification is reduced by cellular material from bioreactor cultivation

We determined the impact of disrupted cells by measuring the amplification efficiency for the GAPDH target using either purified gDNA or suspensions of disrupted cells as template material (Figure 5-4). Disrupted cells from shake flask cultivation had no marked impact on the efficiency profile for the reaction (Figure 5-4 A). By contrast cellular material from bioreactor cultivation constricted the window of efficient amplification from six tenfold dilutions for pure DNA template down to four when disrupted cells are present (Figure 5-4 B).

5.2.5 LRE-qPCR is equivalent to SC qPCR with respect to quantification performance for a CHO genomic target

Two methods of quantitation, the traditional standard curve (SC) qPCR approach and the recently developed method of LRE-qPCR (Robert G Rutledge, 2008), were used for absolute qPCR analysis of copies of the GAPDH sequence within disrupted cell samples. LRE-qPCR was calibrated using the Cal1 primers and methods detailed by Rutledge et al. (2010). Both methods were compared to copy numbers derived from spectrophotometric measurements (dotted lines in Figure 5-6). Four or more tenfold dilutions of the starting material from shake flask cultivation (Figure 5-6 A) should result in samples containing less than one copy of the CHO genome, assuming a genome size of 2.45 Gb. As such it is not unexpected that both SC and LRE-qPCR data diverge from projections based on spectrophotometry after four tenfold dilutions of the initial sample (Figure 5-6 A).

For bioreactor-derived material (Figure 5-6 B), LRE-qPCR data largely agreed with spectrophotometry data over dilutions 2 to 4, after which LRE-qPCR data flattened. There are predicted to be 40 copies of the CHO genome present in the fifth tenfold sample dilution. As such this quantitative limit may be due more to the presence of the disrupted cells than the number of copies of the genome present. Overall, for bioreactor-derived cellular material the LRE-qPCR data matches spectrophotometric projections more closely than SC qPCR over almost every sample dilution (Figure 5-6 B).

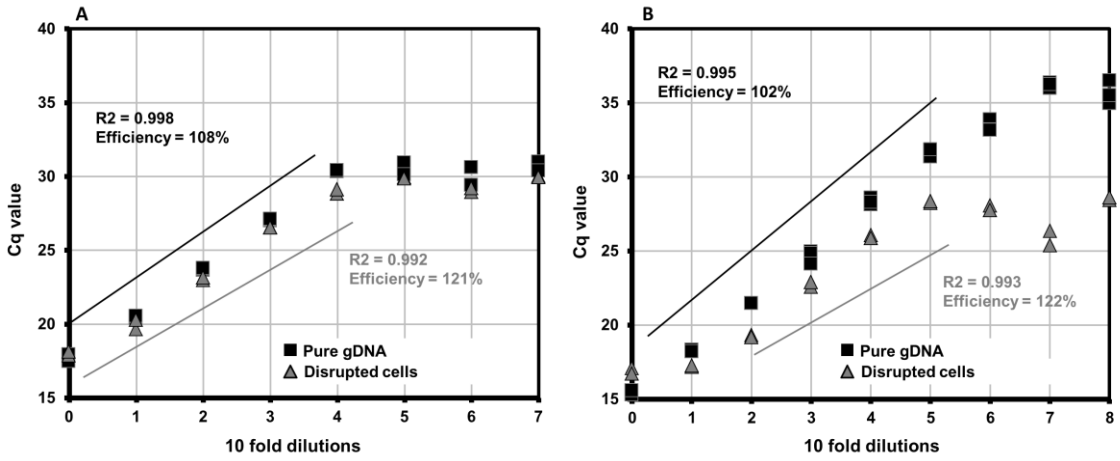

Figure 5-4 Influence of disrupted CHO cells on amplification efficiency for a genomic target. Real time PCR was performed using disrupted cells (grey symbols and lines) or purified DNA (black symbols and lines) from shake flask (A) and bioreactor (B) cultivation as template. The initial shake flask sample contained 2.5x106 cells/mL, which was split into a purified DNA sample and a disrupted cell sample. The bioreactor sample contained 1x1011 cells/mL and was asl split into pure DNA and disrupted cell samples. Cq values were plotted against 10 fold dilutions of template source. Lines indicate data point for which amplification efficiency is 100±10% efficiency, at a R2 > 0.99. Data featured is typical of N=3 analytical repeats.

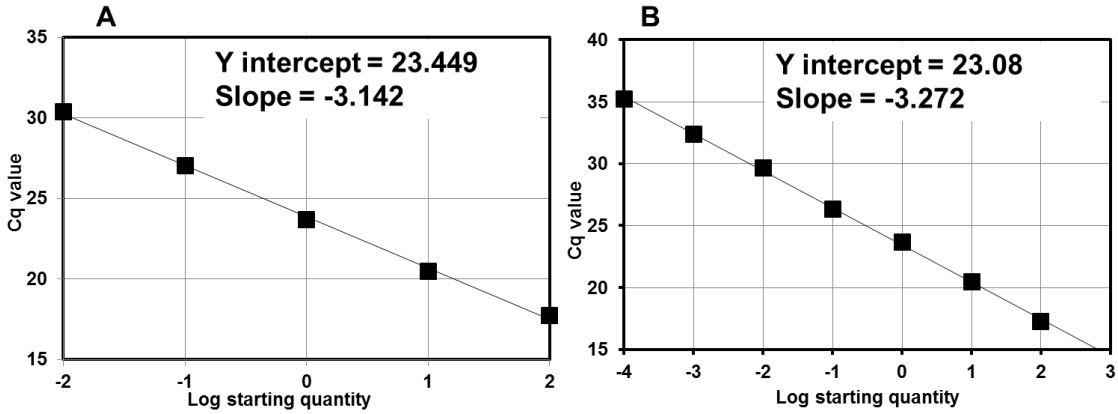

Figure 5-5 Standard curves used in quantification of Cq values generated from sonicated CHO process streams. (A) shows the standard curve generated from purified shake flask material. (B) shows the standard curve generated from purieifed bioreactor material.

Method comparison by XY plot (Burd, 2010) gives a slope of 1.00 and an intercept of zero in the case of zero proportional bias between methods. For shake flask-derived material an XY plot (Figure 5-7 A) showed negligible proportional bias of SC qPCR data (slope of 1.06) when using the LRE-qPCR method. The Y intercept of the XY plot was close to zero (0.0705) indicating little systematic bias. A Bland-Altman (Bland and Altman, 1986) plot of these data (Figure 5-7 B) indicates LRE-qPCR had a positive bias of SC qPCR at higher copy numbers of target DNA but that the methods are broadly equivalent as the mean bias range for both includes zero difference (Burd 10).

For qPCR of bioreactor-derived material an XY plot (Figure 5-7 C) showed that LRE-qPCR exhibited significant proportional (slope of 1.26831) and systemic (intercept of 1.26831) bias of SC qPCR data. However, when each method was separately compared to spectrophotometric data by XY plot their bias of the spectrophotometric data was very similar. This similarity was borne out by Bland-Altman plot (Figure 5-7 D) which indicated LRE-qPCR had negative bias of SC qPCR data at high target DNA copy number but that the methods are statistically equivalent as the bias range spans zero difference.

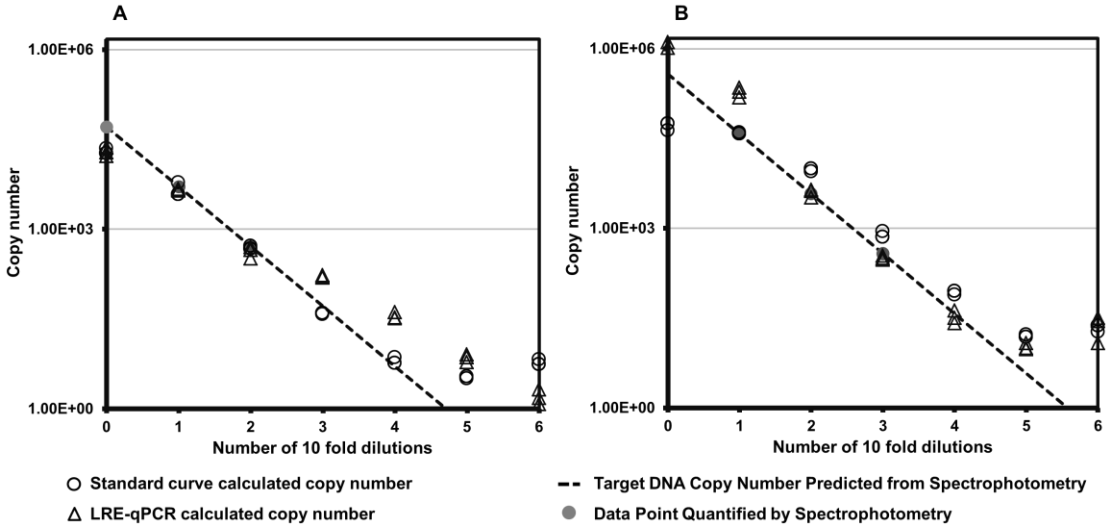

Figure 5-6 Qualitative comparison of SC qPCR and LRE-qPCR for quantitation of a genomic sequence. The predicted number of copies of the GAPDH target sequence in a sample of disrupted cells, as calculated using the LRE-qPCR (open triangles) and SC qPCR (open circles) methods, plotted as a function of sample dilution for samples derived from shake flask (A) and bioreactor (B) cultivation. Grey circles indicate genome copy number inferred from a spectrophotometric measurement of total DNA concentration present in sample. The dashed lines indicate linear extrapolation of the spectrophotometric data points.

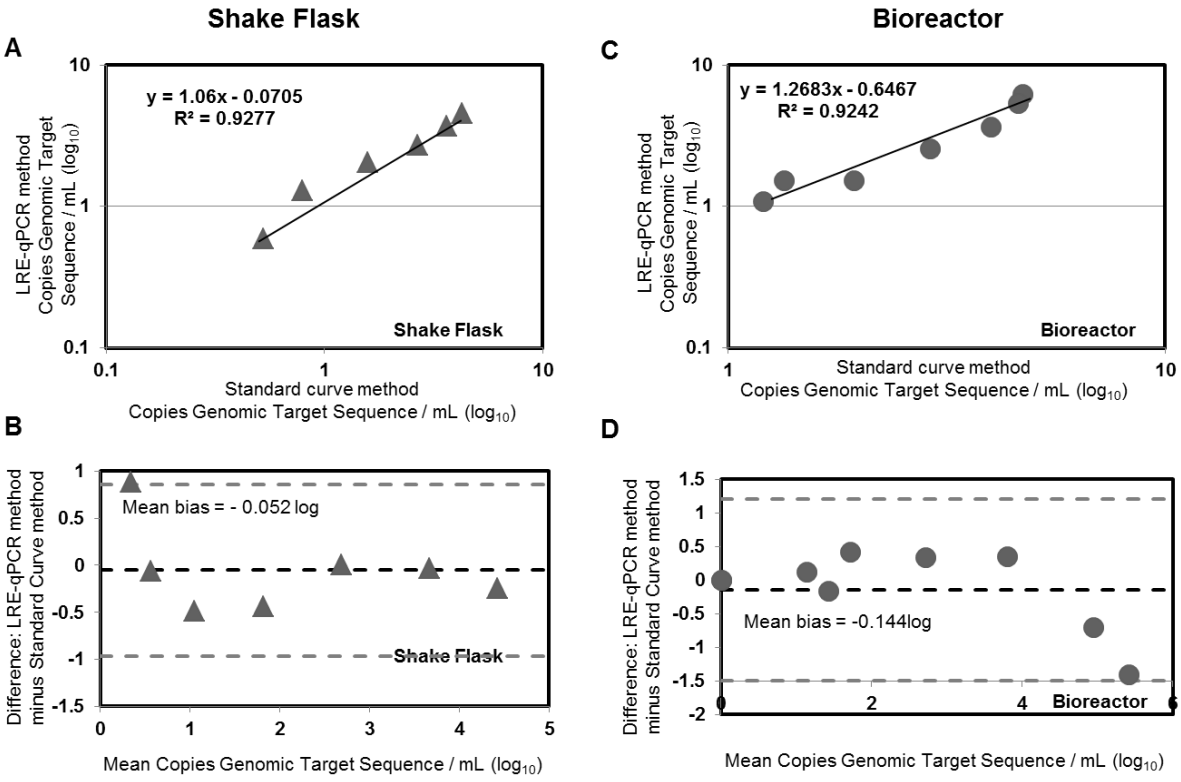

Figure 5-7 Statistical comparison of SC qPCR and LRE-qPCR for quantitation of a CHO genomic sequence. Comparison of SC qPCR and LRE-qPCR methods when used to predicted number of copies of the GAPDH target sequence in a sample of disrupted cells. Data from samples derived from shake flask and bioreactor cultivation was campared using XY plot, graphs A and C respectively, and Bland-Altman plot, graphs B and D respectively. Statistical procedures were performed as described by Burd (2010).

5.2.6 LRE-qPCR quantitation of genomic target is largely unaffected by CHO cellular material

As LRE-qPCR is calibrated from an external lambda DNA sample and quantifies directly from the fluorescence data, we were able to use this method to quantify target in purified DNA samples as well as in disrupted cell solutions. In this way we could evaluate the effect of cellular material on LRE-qPCR performance. For purified DNA samples derived from shake flask cultivation, Figure 5-8 A shows that LRE-qPCR (open squares) agrees well with spectrophotometric data (large dashed line). The presence of disrupted cells from shake flask cultivation caused divergence between LRE-qPCR data (open triangles) and spectrophotometric data (fine dashed line) for the more dilute samples. The equivalent profile for bioreactor-derived samples (Figure 5-8 B) was broadly the same except the presence of disrupted cells caused LRE-qPCR data to diverge from spectrophotometric data at the both the most concentrated and the most dilute samples.

5.2.7 Sensitivity of e-pPCR for mycoplasma DNA sequence detection is depressed by CHO cellular material

A common application of e-pPCR is the binary detection of organisms known to contaminate cultures of mammalian cells at industrial scale. As such we designed a plasmid containing a 300bp sequence conserved across five species of mycoplasma (Kong et al., 2001), as detailed n chapter 3, and used this as a safe proxy test of the sensitivity of e-pPCR for mycoplasma detection. Serial dilutions of the plasmid were made and to each dilution either water or disrupted cells was added (Figure 5-9). Disrupted cells were generated from a sample containing $2x10^6$ cells/mL from shake flask cultivation or from a sample of $2.5x10^5$ cells/mL from bioreactor cultivation (Figure 5-9). The LOD for naked DNA template was 15 copies. This was reduced by one order of magnitude to 150 copies by the presence of disrupted cells, from either shake flask (Figure 5-9 A) or bioreactor (Figure 5-9 B) cultivation. Total amplicon production, with either pure DNA or disrupted cells as template, was similar for shake flask (Figure 5-9 Aii) and bioreactor (Figure 5-9 Bii) cultivation.

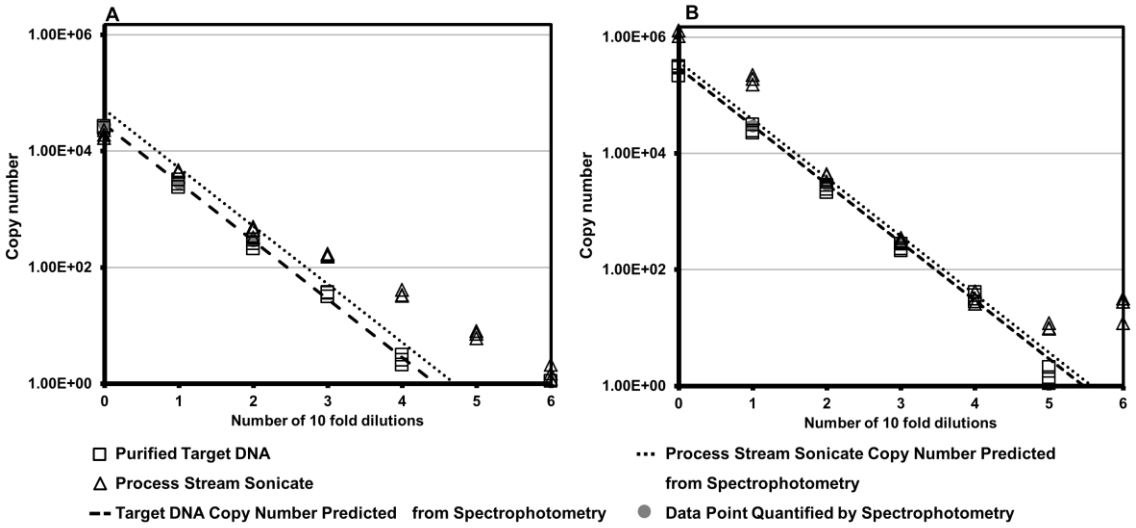

Figure 5-8 Influence of disrupted CHO cells on LRE-qPCR for quantification of a CHO genomic target. The predicted number of copies of the GAPDH target sequence in a sample, as calculated using the LRE-qPCR method, was plotted as a function of sample dilution for samples derived from shake flask (A) and bioreactor (B) cultivation. Samples either underwent total DNA purification (open squares) or only mild cell disruption (open triangles) prior to LRE-qPCR procedure. Genome copy number was also inferred from spectrophotometric measurement of total DNA concentration. These spectrophotometric measurements are indicated by large grey circles and are linearly extrapolated for both purified DNA (thick dashed line) or disrupted cell (fine dashed line) samples.

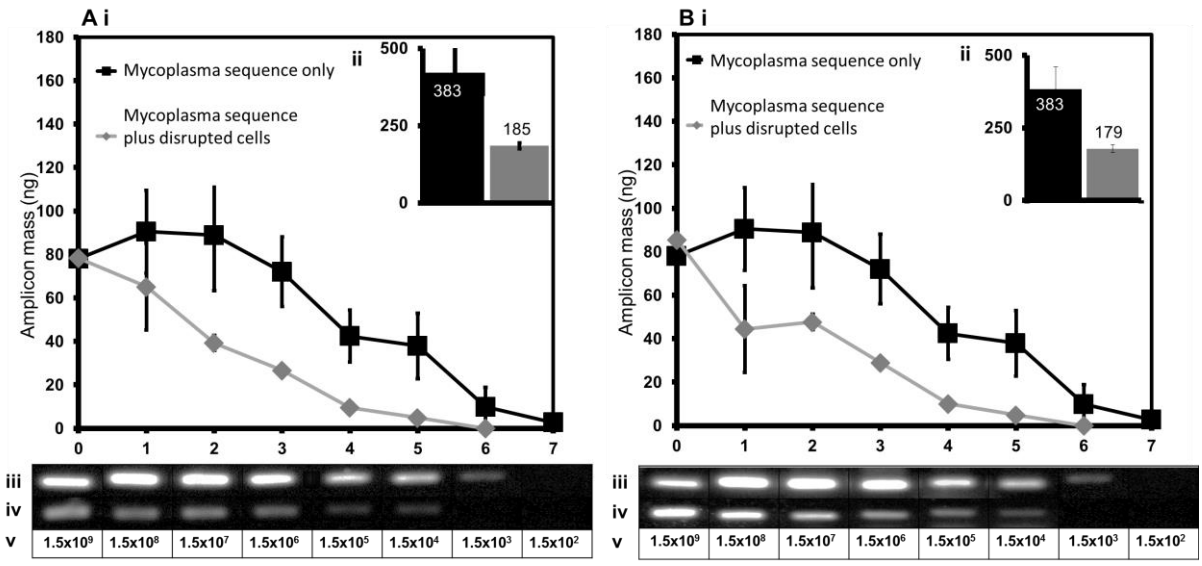

Figure 5-9 Influence of disrupted CHO cells on e-pPCR detection of a mycoplasmal target sequence. 50 ng of plasmid encoding a mycoplasmal DNA sequence was used as e-pPCR template either as purified DNA or purified DNA plus disrupted cells derived from a sample containing X CHO cells from shake flask (panel A) and X CHO cells from bioreactor (panel B) cultivation. For both cultivation methods the following data are depicted. The mass of amplicon produced in a reaction is plotted as a function of sample dilution (i). Inlaid graphs (ii) plot the area under each curve as a bar chart. Agarose gel images show the 184bp amplicons generated from the purified plasmid DNA (iii) and plasmid DNA plus disrupted CHO cells (iv). The number of copies of the plasmid molecule in a given sample is indicated in the row labelled (v).

5.2.8 Amplification efficiency for a mycoplasma DNA sequence largely unaffected by CHO cellular material

A common element of many approaches to absolute qPCR is the importance of amplification efficiency. To evaluate the influence of cellular material on the efficiency of amplification of a mycoplasma sequence we prepared a pure solution of 5ng of plasmid encoding a mycoplasmal sequence and also solution in which 5ng of plasmid was mixed with disrupted cells from shake flask or bioreactor cultivation. A tenfold dilution series of each sample type was then made and Cq values plotted (Figure 5-11). Between two and six tenfold dilutions ($1.5x10^7$ copies to 1500 copies of target sequence) amplification efficiency is largely unaffected by the presence of disrupted CHO cells.

5.2.9 LRE-qPCR and SC qPCR are equivalent with respect to mycoplasma sequence quantification in the presence of disrupted CHO cells

LRE-qPCR (Figure 5-12 A) and SC qPCR (Figure 5-12 B) methods of quantitation were applied to the underlying data used to generate the Cq values in Figure 5-11. Both approaches resulted in reverse S-shaped curves for copy number estimation as a function of template dilution. The dotted line in both Figure 5-12 graphs is an extrapolation of three spectrophotometrically measured data points and serves to aid comparison of LRE-qPCR and SC qPCR data. The two methods are broadly equivalent, with disrupted CHO cells having little effect on copy number estimation over 3-6 tenfold dilutions of template material ($1.5x10^6$ copies to $1.5x10^3$ copies of target sequence).

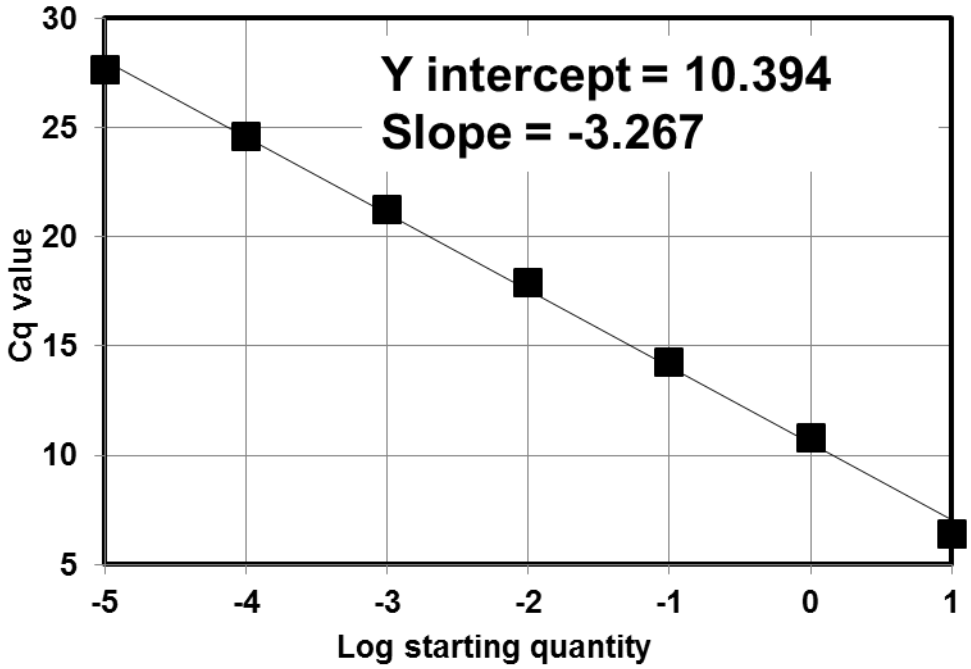

Figure 5-10 Standard curve used in the quantification of Cq values generated from a mycoplasma target

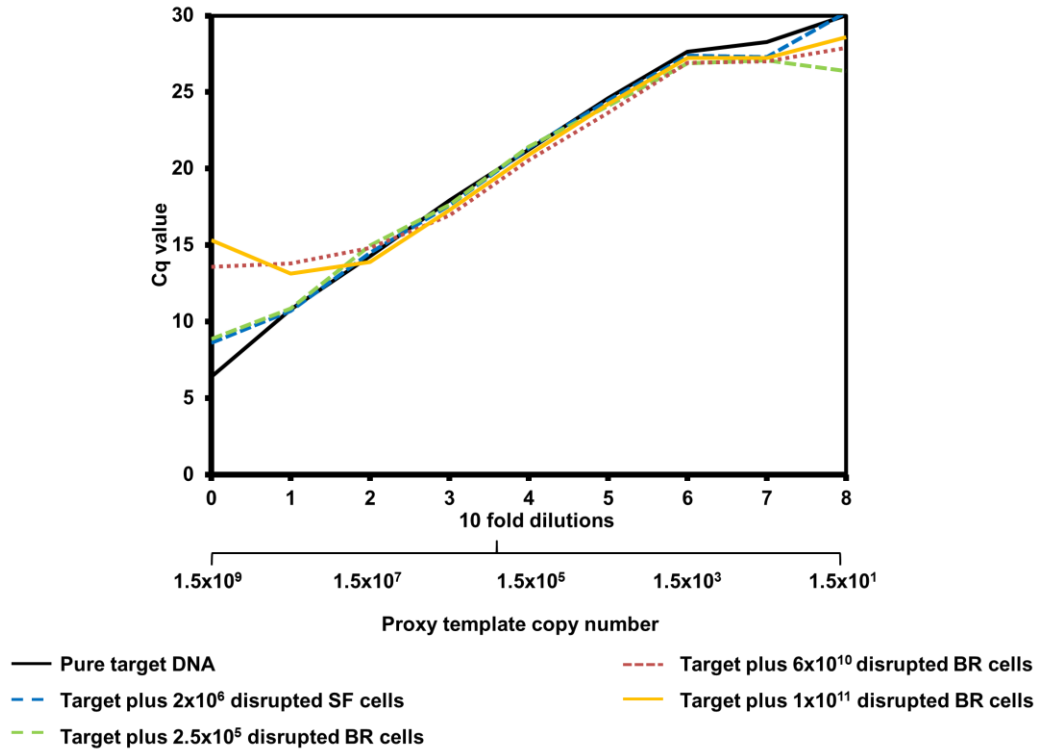

Figure 5-11 Influence of disrupted CHO cells on amplification efficiency for a mycoplasmal target sequence. Real time PCR was performed using 50 ng of plasmid encoding a mycoplasmal DNA sequence and a further 8 tenfold dilutions of the plasmid template. For each starting solution either zero cells were present or disrupted cells derived from samples of $2x10^6$ cells/mL, from shake flask cultivation, or $2.5x10^5$ cells/mL, $6x10^{10}$ cells/mL and $1x10^{11}$cells/mL, from bioreactor cultivation, were added to plasmid DNA (see legend). The resultant Cq values for each amplification reaction were plotted as a function of sample dilution. Data featured is typical of N=3 analytical repeats.

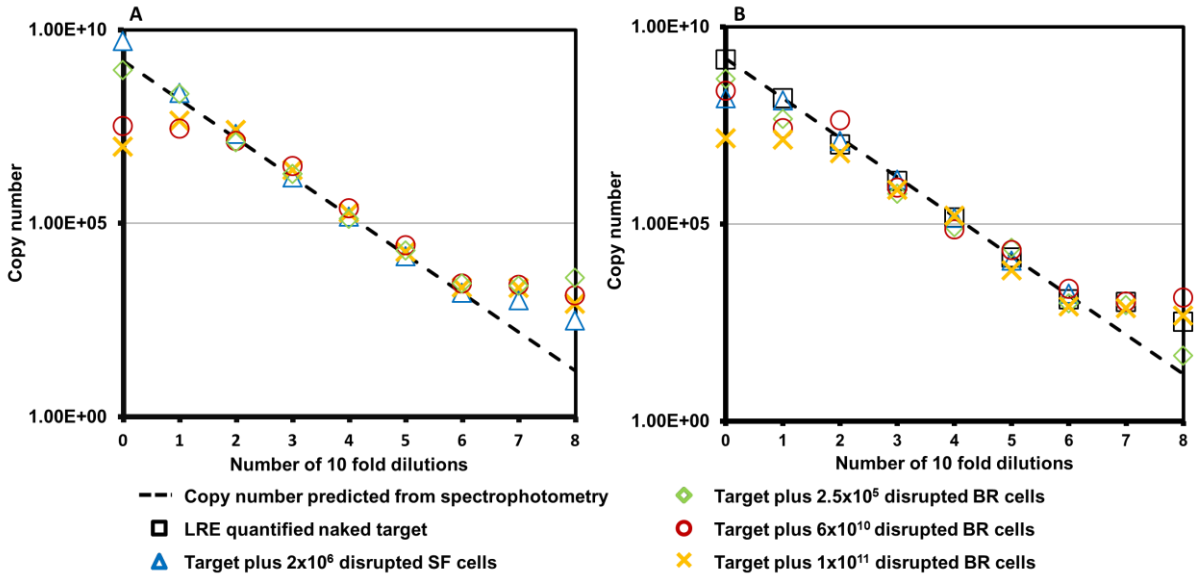

Figure 5-12 Comparison of SC qPCR and LRE-qPCR methods for absolute quantitation of a mycoplasmal DNA sequence. SC qPCR (A) and LRE-qPCR (B) methods were used to quantify the number of copies of a mycoplasmal DNA sequence present in reactions containing plasmid DNA plus disrupted CHO cells derived from a 2x106 cells/mL shake flask sample or 2.5x105 cells/mL, 6x1010 cells/mL and 1x1011cells/mL bioreactor cultivation samples (see legend). The number of copies of the plasmid in a given sample was also inferred from spectrophotometric measurement of total DNA concentration before addition of disrupted cells. These spectrophotometric measurements are indicated by large grey circles and are also linearly regressed (thick dashed line) for qualitative comparison. For LRE-qPCR (B) it is possible to assess quantification of the pure mycoplasmal sequence (open squares) because the unrelated Cal1-OCF reaction is used for calibration. For SC qPCR (A), quantification of the pure mycoplasmal sequence is not informative as this reaction represents the standard curve used for calibration.

5.3 Discussion

As the number of innovative approaches to mammalian cell genome and gene network implementation expands so the need for standards in synthetic biology also becomes more acute. Industrial application of synthetic biology also requires standards that enable regulatory compliance and accurate analysis of chassis and bioprocess performance. The purpose of PCR sample preparation is to remove inhibitors that could lead to false positives, false negatives or inaccurate quantification. Sample preparation procedures tend to significantly extend assay duration so that live or at-line data set capture is not possible. This in turn delimits application of PCR approaches in statistical process optimisation procedures such as design of experiments (DOE). To address these issues of standardisation and sample preparation we sought to test the following hypotheses for both e-pPCR and qPCR: i) that the removal of cellular material may not necessarily be required for certain PCR-based assays and ii) that the LRE-qPCR method, which incorporates a putatively universal standard which is therefore potentially of use to synthetic biologists, is equivalent to conventional SC qPCR in terms of sensitivity and accuracy.

We used the MIQE-compliant primers designed in chapter 2 to amplify a sequence present as a single copy within the CHO genome and a mycoplasmal sequence common to many mycoplasma species known to infect mammalian calls. We also used the ‘Cal1’ primers that comprise the OCF1 calibration reaction for the LRE-qPCR method. We then cultivated CHO cells in shake flasks and in an industrially relevant rocked bag bioreactor (Figure 5-1). We determined that the LOD for e-pPCR when used to detect a genomic target (Figure 5-3) and a mycoplasmal sequence (Figure 5-9) is reduced tenfold by the presence of disrupted CHO cells. This indicates clearly that sample preparation is required for accurate and sensitive use of e-pPCR as a detection method.

Unlike e-pPCR, qPCR data collection occurs during the reaction, thus making any sample preparation time a larger fraction of total assay throughput time. We determined the extent to which the presence of disrupted CHO cells affects amplification efficiency

- a key metric for multiple statistical approaches to analysis of quantitative real-time PCR methods. The presence of cellular material from shake flasks had no effect on genomic target amplification efficiency (Figure 5-4 A) whereas material from bioreactors did constrict the range of reactions for which acceptable amplification efficiency was observed (Figure 5-4 B). This indicates that, for multiple qPCR methods, only minimal and rapid sample preparation is required for samples taken in the early, seed train, stages of industrial CHO cell cultivation. For amplification of a mycoplasmal sequence (Figure 5-11), the presence of CHO material originating from shake flask or bioreactor cultivation influenced amplification efficiency only in very concentrated or very dilute samples.

LRE-qPCR, as reported by Rutledge et al. (2010), features the CAL1 reaction for calibration, which consists of lambda bacteriophage genome as target and a high-performance 'Cal1' primer pair. By contrast most academic and industrial organisations use unique qPCR primer sets and, in the case of absolute quantitation, construct their own standard curves. Quantification from a standardised source could introduce far greater reproducibility of absolute qPCR data across facilities. As such we suggest adoption of Cal1/OCF reaction as a qPCR standard represents an excellent opportunity to improve standardisation within the synthetic biology and biotechnology communities. Standardisation between qPCR assays currently extends only to experimental setups, information reporting, such as the MIQE guidelines, and testing of food and water sources for contaminants.

We compared the accuracy of LRE-qPCR and conventional SC qPCR for quantification of a genomic target sequence by juxtaposing both methods with spectrophotometric data (Figure 5-6) and by statistical head-to-head analysis (Figure 5-7). LRE-qPCR matched the performance of conventional SC qPCR for this target. The equivalence of LRE-qPCR and SC qPCR could also be seen for quantification of a mycoplasmal sequence, comparing each method to spectrophotometric data (Figure 5-12). This was the case both in the presence and absence of disrupted CHO cells. Critically, disrupted CHO cells had very little effect on the accuracy of LRE-qPCR for quantitation of a genomic

target compared to spectrophotometric data (Figure 5-6 and Figure 5-8) even when sourced from bioreactor cultivation at high cell concentrations ($1x10^{11}$ cells / mL).

In conclusion, we suggest that sample preparation is necessary for e-pPCR as a detection tool for CHO cell and mycoplasmal DNA.

5.4 Conclusions

- Sample preparation is necessary for e-pPCR as a detection tool for CHO cell and mycoplasmal DNA.
- Cellular material from shake flasks had no effect on genomic target amplification efficiency but material from bioreactors limited the range of reactions for which acceptable amplification efficiency was observed - indicates that, for multiple qPCR methods, only minimal and rapid sample preparation is required for samples taken in the early, seed train, stages of industrial CHO cell cultivation.
- For amplification of a mycoplasmal sequence the presence of CHO material originating from shake flask or bioreactor cultivation influenced amplification efficiency only in very concentrated or very dilute samples.
- LRE-qPCR matched the quantitative accuracy of SC qPCR

Chapter 6

6 Influence of high cell density *Escherichia coli* cells on performance of PCR as a synthetic biology tool for bioprocess monitoring and contaminant detection

6.1 Introduction

6.1.1 Synthetic prokaryotic genomes and standards for their quantification

As with eukaryotic genomes, synthetic biology has ushered in a new era in which entirely synthetic (Gibson et al., 2010) or refactored (Lajoie et al., 2013) genomes can be used to control bacterial cells. As this field of 'synthetic genomics' progresses it is inevitable that ever more *de novo* genome design strategies will be conceived and tested by synthetic biologists in both basic research and industrial settings. The current trajectory of much synthetic biology research make it increasingly likely that quantitative analysis of host cell genetic barcode (Parodi et al., 2002), genetic drift (Voronin et al., 2009), genome-integrated transgene copy number (Cong et al., 2013), plasmid copy number (Lee et al., 2006b) and natural (Katariina E. S. Tolvanen, 2008) or synthetic consortia (Bernstein and Carlson, 2012) will become critical factors for industrial application. We believe the integrity of large and complex gene networks present within genomes and refactored genomes will all drive greater need for at-line, quantitative analysis of genomic loci during industrial procedures.

Bacteriophage can compromise virtually any industrial process involving bacteria (Sturino and Klaenhammer, 2006), even those with comprehensively refactored genomes (Lajoie et al., 2013). In this study we test PCR approaches to detecting bacteriophage DNA in industrial process streams. We have used a proprietary primer set specific to a 300bp region of bacteriophage DNA encoded by a plasmid (pPROX1) as a proxy for actual detection of bacteriophage particles in process streams. The use of a proxy sequence in this way enables investigation of detection of many pathogen types without the need to risk infection of other cultivation experiments being performed in the same facility.

Our aims in this chapter are to firstly measure the extent to which sample preparation is necessary for PCR analysis of target sequences and then rigorously compare the performance of LRE-qPCR and SC qPCR methods. These investigations measure i) material from shake flask and high cell density bioreactor cultivation and ii) measuring a genomic locus and a plasmid-encoded bacteriophage sequence as a target sequences. Finally will also discuss the suitability of LRE-qPCR as a synthetic biology standard for the qPCR of prokaryotic organisms.

6.2 Results

6.2.1 PCR design and cell cultivation

We used shake flasks to grow an *E. coli* W3110 production strain in defined media (Balasundaram et al., 2009) to OD_{600} = 2.5 (Figure 6-1 A), typical of the end-point of seed train cultivation used to provide inoculum for growth in bioreactors. The strain harbours the plasmid pTTOD-A33 which encodes a recombinant Fab' fragment (Nesbeth et al., 2012) inducible by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG). Bioreactor cultivation started in batch mode (Figure 6-1 B) to OD_{600}=130 then fed-batch mode was applied along with IPTG addition to induce Fab' fragment expression. An experimental sample was taken two hours post-induction, at OD_{600}=160, during early idiophase growth when bacteriophage contamination can be highly costly.

The proprietary pPROX1 plasmid encoding 300bp of bacteriophage DNA was used as a proxy for bacteriophage particle detection and quantitation. After sample processing, pPROX1 was added at known concentration and the ability of PCR methods to detect or quantify the bacteriophage sequence was tested.

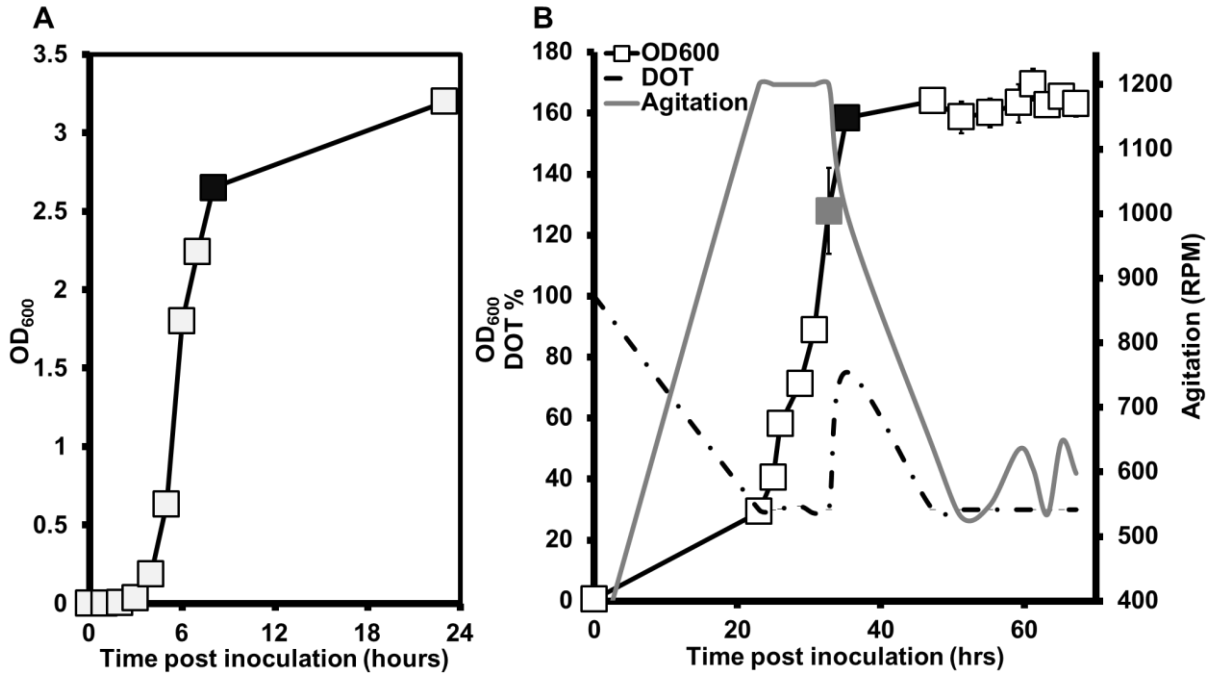

Figure 6-1 Shake flask and bioreactor cultivation of a W3110 E. coli production strain. 40 mL of on E. coli W3110 production strain grown in LB was used to inoculate 360 mL defined media in a 5L shake-flask (A). An uninduced sample was taken at the start of stationary phase growth in shake flasks (black filled square) for PCR experiments. 10% of this culture was used to inoculate 3.6 L defined media in a New Brunswick 7L bioreactor (B). In bioreactor cultivation, IPTG was added to induce transgene expression at 34 hours post-inoculation (grey filled square) and a sample taken 2 hours post-induction (black filled square) for PCR experiments. Error bars represent standard deviation over n=3 experimental repeats.

6.2.2 *E. coli* cellular material reduces ability of e-pPCR to detect genomic target

We determined the extent to which the presence of disrupted *E. coli* cells influences the LOD of e-pPCR for a single copy genomic *E. coli* target sequence. We took samples from shake flask and bioreactor cultivation (Figure 6-1) and split each sample. For one half of the sample, we purified total nucleic acids by phenol chloroform method and used this as purified target DNA in subsequent experiments. The other half-sample was subjected to a brief sonication procedure in order to liberate host DNA from cells. Gel electrophoresis of this sample before and after sonication revealed that no discernible shearing of genomic DNA had occurred (Figure 6-2).

We used agarose gel analysis and densitometry to match purified DNA samples and disrupted cell samples with respect to their host DNA content. We then set up a series of tenfold sample dilutions and defined LOD as the first tenfold dilution for which amplicon band was undetectable after n=3 repeats. Figure 6-3 A shows e-pPCR LOD of 0.5pg of pure gDNA, which is reduced to 5pg gDNA when disrupted cells from shake flask cultivation are present. The same LOD profile was observed for bioreactor-derived material (Figure 6-3 B).

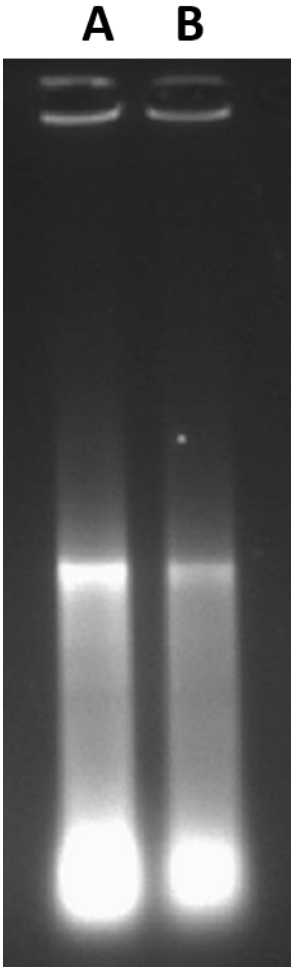

Figure 6-2 Comparison of nucleic acid extracted (A) and sonicated (B) *E. coli* gDNA integrity

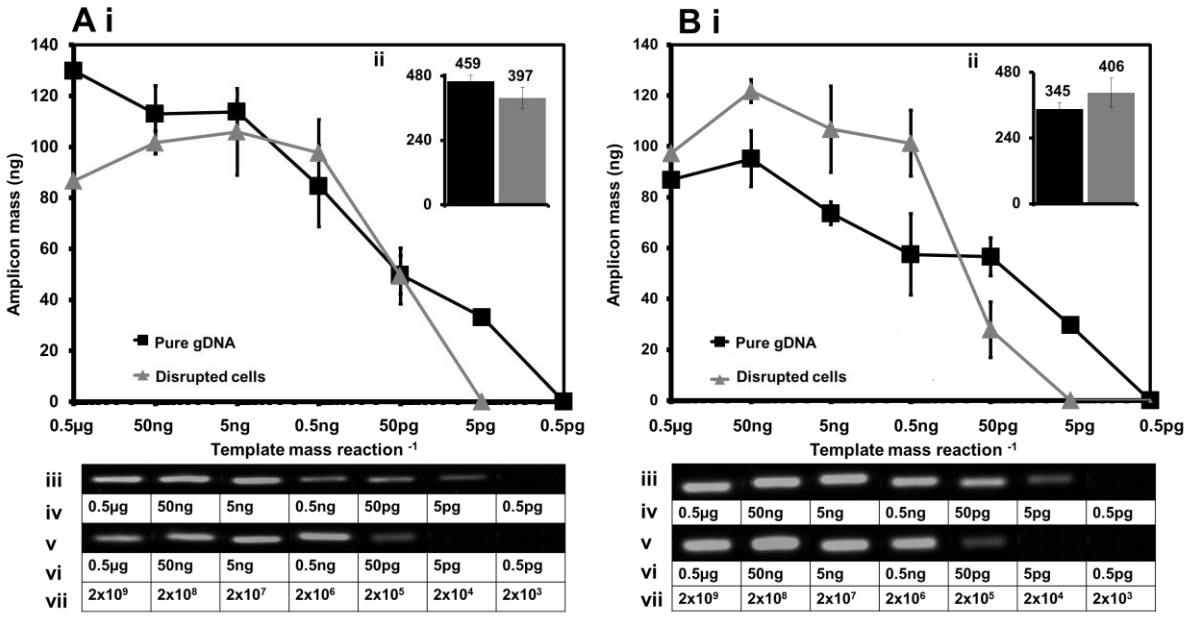

Figure 6-3 Influence of disrupted E. coli cells on e-pPCR for detection of a genomic target sequence. Purified DNA (black symbols and lines) or disrupted cells (grey symbols and lines) was used as template for e-pPCR and final amplicon yield measured by densitometric analysis of the resultant band from gel electrophoresis. Samples were from either shake flask (Panel A) or bioreactor (Panel B) E. coli cultivation. The following sections are present in both Panels A and B. Section i is a plot of densitometry measurements of the resultant 313bp amplicon band as a function of tenfold dilutions of sample. Inlaid graph (Section ii) plots the area under each curve in Section A, with the total value indicated above the bar. Agarose gel photos (Section iii) and estimated template mass (Section iv) are indicated for pure DNA template reactions. For reactions in which disrupted cells provide PCR template, amplicon gel photos (Section v), estimated template mass (Section vi) and cell numbers present pre-disruption (Section vii) are indicated. Error bars represent standard deviation over n=3 experimental repeats.

6.2.3 Cellular material from HCD bioreactor cultivation reduces qPCR efficiency

To evaluate the degree to which cellular material influences qPCR performance we set up reactions in the presence and absence of disrupted cells. Efficiency of amplicon production, defined as the slope of Cq values when plotted as a function of reaction cycle, is a key element for many of the numerous statistical approaches to qPCR data analysis. Typically, efficiency within the range of 100±10%, with R^2 of at least 0.99, is set as the limit for accurate quantitation.

Shake flask material (Figure 6-4 A) from which DNA was purified, resulted in a 'window' of data points with acceptable efficiency that spanned 3 tenfold dilutions. The presence of disrupted cells from shake flask cultivation did not widen or constrict this window of operation.

For bioreactor material (Figure 6-4 B), pure DNA samples were amplified with 100±10% efficiency over 5 tenfold dilutions. Disrupted cells from bioreactor cultivation constricted this window of operation to 3 tenfold dilutions. For the undiluted bioreactor material, the disrupted cell sample and purified DNA sample contained 5µg/µL and 3.9µg/µL respectively. These high DNA concentrations inhibited amplification for the disrupted cell sample and abolished amplification in the purified DNA sample, which we believe was due to inhibition of polymerase activity.

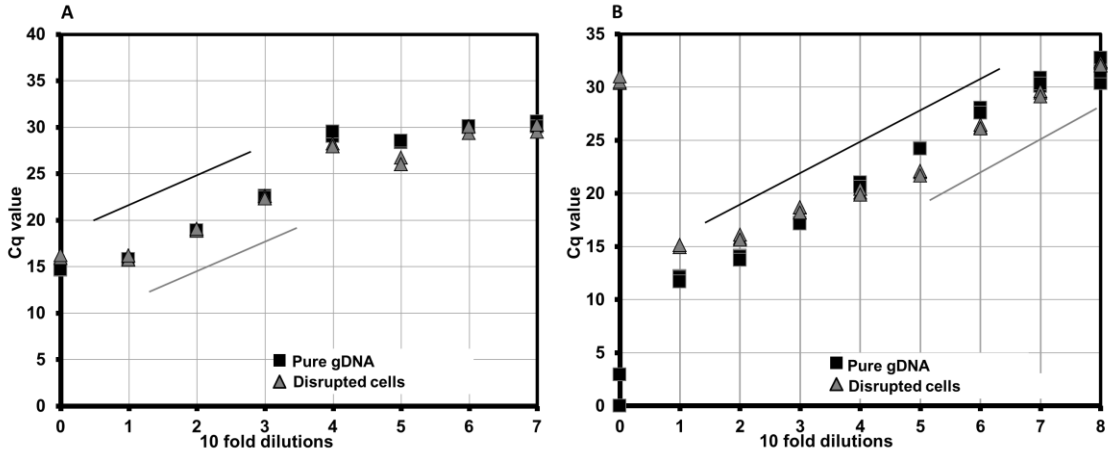

Figure 6-4 Influence of disrupted *E. coli* cells on amplification efficiency for a genomic target sequence. Real time PCR was performed using template material either from process stream sonicates (grey symbols and lines) or purified DNA (black symbols and lines) from shake flask (A) and bioreactor (B) cultivation. Spectrophotometry indicated the undiluted, purified DNA sample derived from shake flask material contained 112ng DNA and the undiluted total process stream sample contained 215ng. For bioreactor material, 3.9µg and 5µg respectively of DNA was present in purified and total process stream samples. Lines indicate data point for which amplification efficiency is 100±10% efficiency, at a R2 > 0.99. N=3 analytical repeats.

6.2.4 LRE-qPCR and SC qPCR are equivalent for genomic target quantification

Spectrophotometry has been used by others to assess the accuracy of qPCR (Shokere et al. 2009), using purified DNA as template. Unexpectedly, we observed in this study that spectrophotometry could be used to measure DNA concentration in the presence of disrupted cells from samples of up to OD_{600}=16 (Figure 6-6), even though we anticipated the presence of such cell debris would distort the absorbance spectra. As such we used spectrophotometry as a mechanistically distinct comparator method to enable qualitative comparison of the performance of LRE-qPCR and SC qPCR.

The genomic target sequence is present in the *BirA* gene, which is known to be present as a single copy in the *E. coli* genome. As such we assumed that total *BirA* copies are equivalent to total genome copies. Assuming an *E. coli* W3110 genome size of 4,646,332bp plus pTTOD-A33 plasmid size of 6480bp with copy number of 10-20 replicons / cell, host gDNA should constitute 97.2%-98.6% of total host cell DNA in process streams. Given these figures, we could convert DNA concentration levels, derived from spectrophotometry, to target sequence copy numbers.

Three spectrophotometric measurements were taken, plotted (Figure 6-6 A and Figure 6-6 B, grey circles) and extrapolated to provide a common benchmark for comparison of LRE-qPCR and SC qPCR. For quantitation of target DNA within disrupted cell samples from shake flask growth (Figure 6-6 A), LRE-qPCR data points matched the trend of spectrophotometric data more closely than SC qPCR. This observation also holds for high cell density bioreactor material (Figure 6-6 B) with the caveat that both LRE-qPCR and SC qPCR were unable to quantitate target DNA in undiluted samples and diverged significantly from the spectrophotometric for the first two tenfold dilutions.

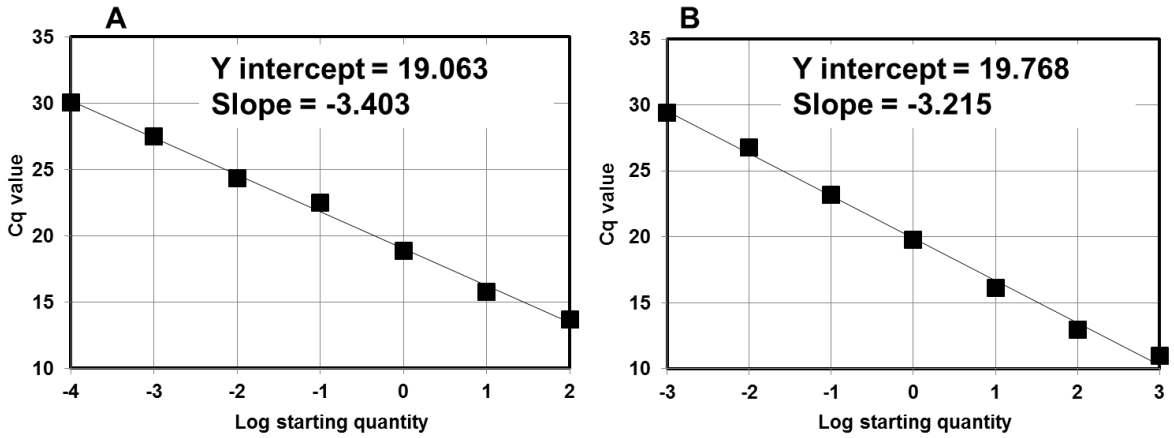

Figure 6-5 Standard curves used in quantification of Cq values generated from sonicated *E. coli* process streams. (A) shows the standard curve generated from purified shake flask material. (B) shows the standard curve generated from purieifed bioreactor material.

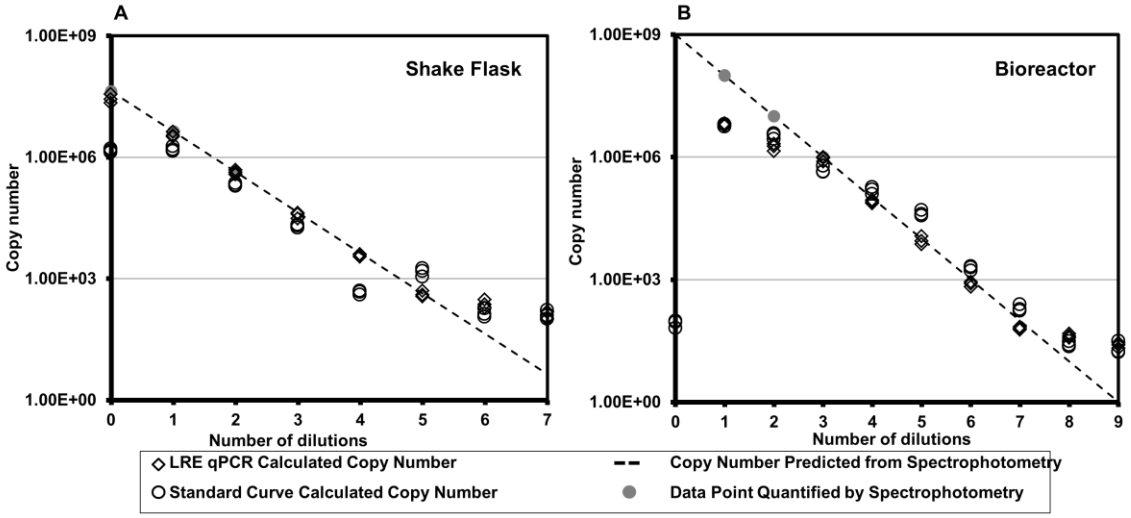

Figure 6-6 Absolute quantification of an *E. coli* genomic target sequence using spectrophotometry, SC qPCR and LRE-qPCR. Grey data points indicate spectrophotometric data and dashed lines extrapolate these data to predict copy number at lower dilutions of both shake flask (graph A) and bioreactor (graph B) samples. Circles indicate copy number determined by the Standard Curve method, rhomboids indicate copy number determined by LRE method.

Method comparison by XY plot (Burd 10) gives a slope of 1.00 in the case of zero bias between methods. When compared to SC qPCR using an XY plot (Figure 6-7 A), LRE-qPCR showed a small degree of proportional bias (slope of 1.1351) for quantitation of target in disrupted cells from shake flask cultivation. The Y intercept of 0.16674 also suggests modest systematic bias. A Bland-Altman (Bland and Altman, 1986) plot of these data (Figure 6-7 B) indicates LRE-qPCR had a slightly negative bias of SC qPCR data but that the methods are equivalent due to the fact that the mean bias range includes zero difference (Burd 10). Comparison of LRE-qPCR and SC qPCR analysis of disrupted cells from bioreactor cultivation (Figure 6C) showed less proportional bias (slope of 1.0074) than for shake flask material but greater systemic bias (Y intercept of 0.2053). Bland-Altman plot again indicated the methods are equivalent as the mean bias range spans the zero difference level (Burd, 2010).

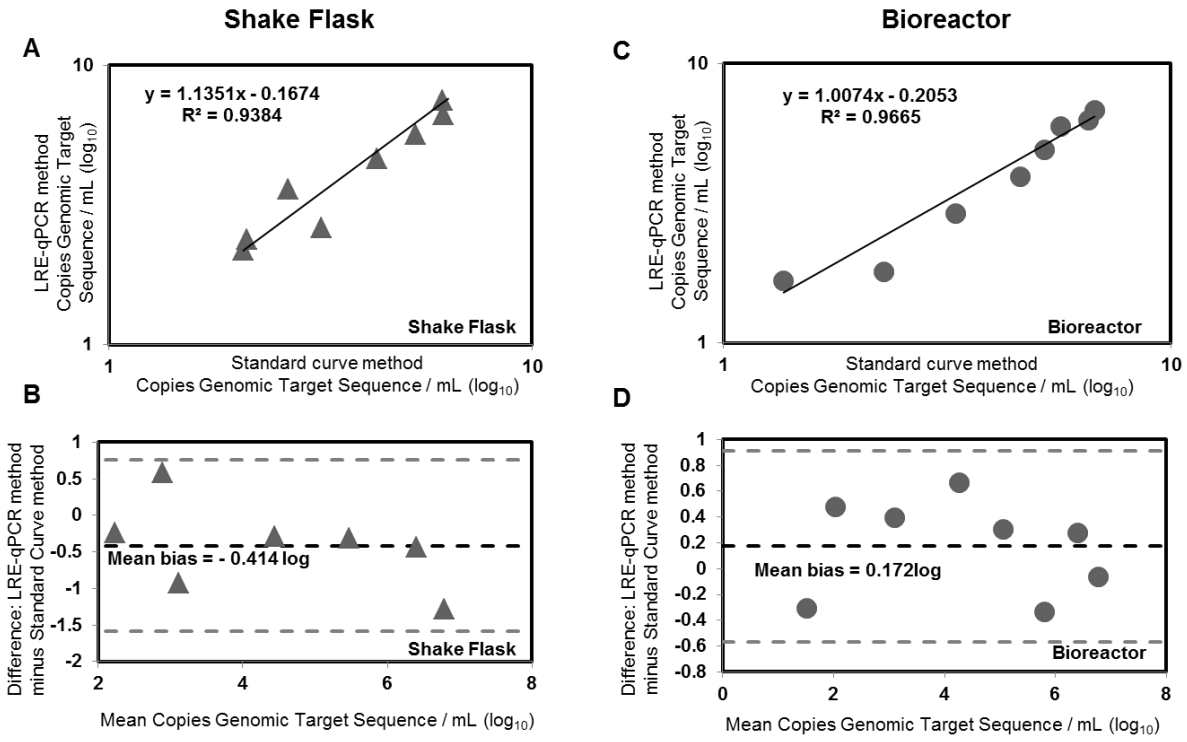

Figure 6-7 Statistical comparison of SC qPCR and LRE-qPCR for quantification of an E. coli genomic target sequence. XY plots (graphs A and C) were derived from copy number estimations resulting from each indicated method, plotted in Figure 5. Bland-Altman bias plots (graphs B and D) were derived from XY plot analyses. Statistical procedures were performed as described by Burd (2010).

6.2.5 Cellular material has minimal effect on LRE-qPCR performance

SC qPCR is predicated on the use of a standard curve comprised of the same primers and target as those used in experimental reactions. Because purified DNA was used as standard curve for the SC qPCR experiment in Figure 6-6 it cannot meaningfully be used to evaluate the accuracy of SC qPCR for purified DNA template. No such restriction applies to LRE-qPCR and as such we plotted LRE-qPCR data gathered using naked DNA and disrupted cell suspension as template alongside data points generated by spectrophotometry (Figure 6-8).

For disrupted cell samples derived from shake flask cultivation (Figure 6-8 A), there is close agreement between LRE-qPCR and spectrophotometric data, for undiluted material and over five tenfold dilutions. The copy numbers indicated by LRE-qPCR plateau over tenfold dilutions 5-7, suggesting either a false positive or that a certain level of DNA remains permanently associated with cellular material.

For samples derived from bioreactor cultivation (Figure 6-8 B) the presence of disrupted cells had virtually no effect on LRE-qPCR measurements. At high concentrations, both purified DNA and disrupted cells depressed the target numbers indicated by LRE-qPCR compared to spectrophotometry. This is likely due to the presence of up to 5µg of DNA, which would be expected to inhibit amplicon production. Overall, LRE-qPCR closely matched spectrophotometric prediction of copy number (Figure 6-8 B) over tenfold dilutions 3-8 for purified target DNA and 3-7 for process stream material.

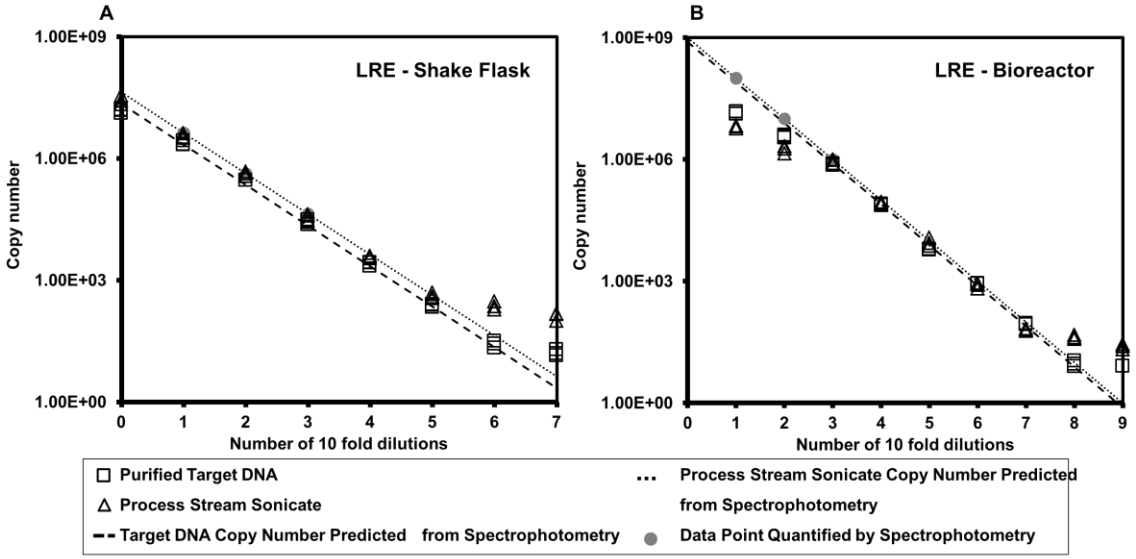

Figure 6-8 Influence of disrupted E. coli cells on LRE-qPCR quantification of an E. coli genomic target sequence. The LRE method was applied to real time PCR fluorescence data gathered using dilutions of shake flask (A) and bioreactor material (B) and pure gDNA extracted from these materials. Grey data points indicate spectrophotometric data and dashed lines extrapolate these data to predict copy number at lower dilutions.

6.2.6 *E. coli* cellular material reduces e-pPCR sensitivity to bacteriophage target

We next evaluated the effect of cellular material on bacteriophage DNA sequence detection by e-pPCR (Figure 6-9). We performed e-pPCR using proprietary primers and serial dilutions of naked pPROX1 plasmid as target. We repeated the reactions in the presence of a constant volume of disrupted cells and defined LOD as the first tenfold dilution of template that yielded no amplicon. For purified DNA, an e-pPCR LOD of 50 copies of the bacteriophage target sequence was observed (Figure 6-9 Ai). The presence of disrupted cells from a shake flask culture sample of OD_{600}=2.5 reduced this LOD tenfold to 500 copies (Figure 6-9 Ai). Disrupted cells from OD_{600}=5 bioreactor culture reduced the e-pPCR LOD a hundredfold to 5000 copies (Figure 6-9 Bi).

6.2.7 Minimal effect of cellular material on qPCR of bacteriophage target sequence

We performed real time PCR with 50ng of the naked pPROX1 plasmid, encoding bacteriophage DNA sequence, as template and plotted Cq values as a function of tenfold dilutions (Figure 6-11) to assess amplification efficiency. For naked template DNA, efficiency within the range of 100±10%, with a $R^2 \geq 0.99$, was observed across tenfold dilutions 0-8. The lowest level of naked template DNA, 140 copies in the ninth tenfold dilution, gave the same Cq value as 1400 copies. The 'window of efficiency' observed for naked DNA extended from the undiluted sample, containing 50ng of plasmid and over 8 subsequent tenfold dilutions (Figure 6-11).

We next made up samples in which 50ng of the naked pPROX1 plasmid was added to cell debris derived from OD_{600}=2.5 shake flask material and bioreactor material of OD_{600}=5, OD_{600}=50 and OD_{600}=160. These starting samples were then serially diluted and used as template for qPCR. Disrupted cells from OD_{600}=2.5, OD_{600}=5 and OD_{600}=50 cultures required two tenfold dilutions to achieve efficient amplification (Figure 6-11). Three tenfold dilutions were required when OD_{600}=160 was present. When disrupted material from any of the four cell concentrations was present, copy number estimation would also flatten out after 7-8 tenfold dilutions, predicted to contain 150 and 15 copies of the target sequence respectively.

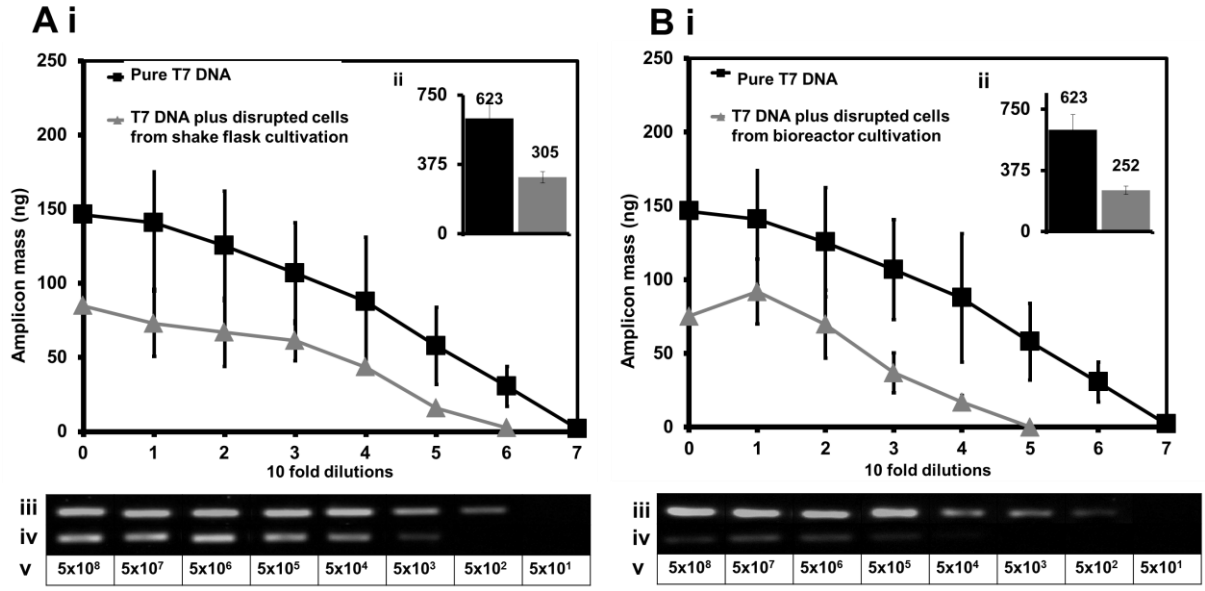

Figure 6-9 Influence of disrupted E. coli cells on e-pPCR quantification of a bacteriophage target sequence. A series of tenfold dilutions of plasmid-encoded bacteriophage target sequence was used as e-pPCR template either as naked DNA in water (black symbols and lines) or in the presence of process stream material (grey symbols and lines) of OD600=2.5 from shake flask (panel X) or OD600=5 from bioreactor (panel Y). The following sections are present in both panels A and B. Section i is a plot of densitometry measurements of the resultant 242bp amplicon band as a function of tenfold dilutions of template DNA. Inlaid graph (Section ii) plots the area under each curve in Section i, with the total value indicated above the bar. Agarose gel photos (Section iii) are of amplicon bands from pure DNA template reactions or those in with process stream material is present (Section iv), with predicted template copy numbers indicated in Section v.#

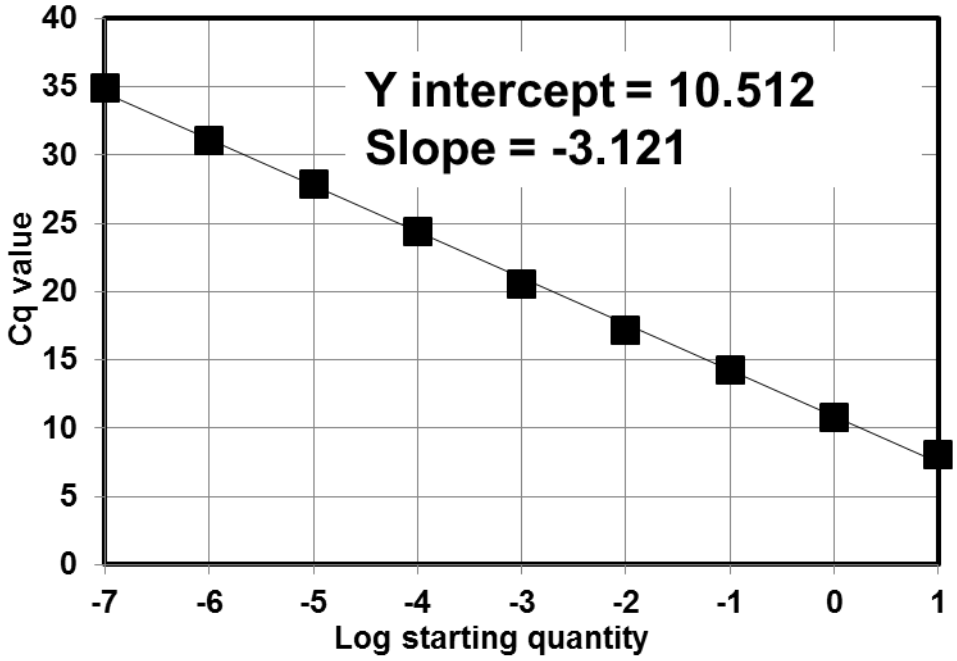

Figure 6-10 Standard curve used in the quantification of Cq values generated from a T7 phage target

6.2.8 3.8 LRE-qPCR outperforms SC qPCR for quantification of bacteriophage DNA in presence of high levels of cellular material

We next used LRE-qPCR and SC qPCR to derive absolute bacteriophage DNA copy numbers from the real time PCR experiments described above in discussion of Figure 9. For comparison, absolute copy numbers calculated by the two different methods were plotted alongside copy numbers derived from spectrophotometry of naked plasmid DNA (Figure 6-12). For SC qPCR (Figure 6-12 A), both the high and low extremes of bacteriophage DNA concentration resulted in divergence from spectrophotometric data. The LRE-qPCR method provides copy numbers in agreement with spectrophotometric data in the range of $1.5x10^4$-$1.5x10^9$ copies of target, despite the presence of process stream material of OD_{600} up to 160 in the undiluted sample. Over all, LRE-qPCR determinations of copy number agree better with spectrophotometric copy numbers at high concentration of process material and bacteriophage DNA but less so at much lower bacteriophage DNA concentrations (Figure 6-12 B).

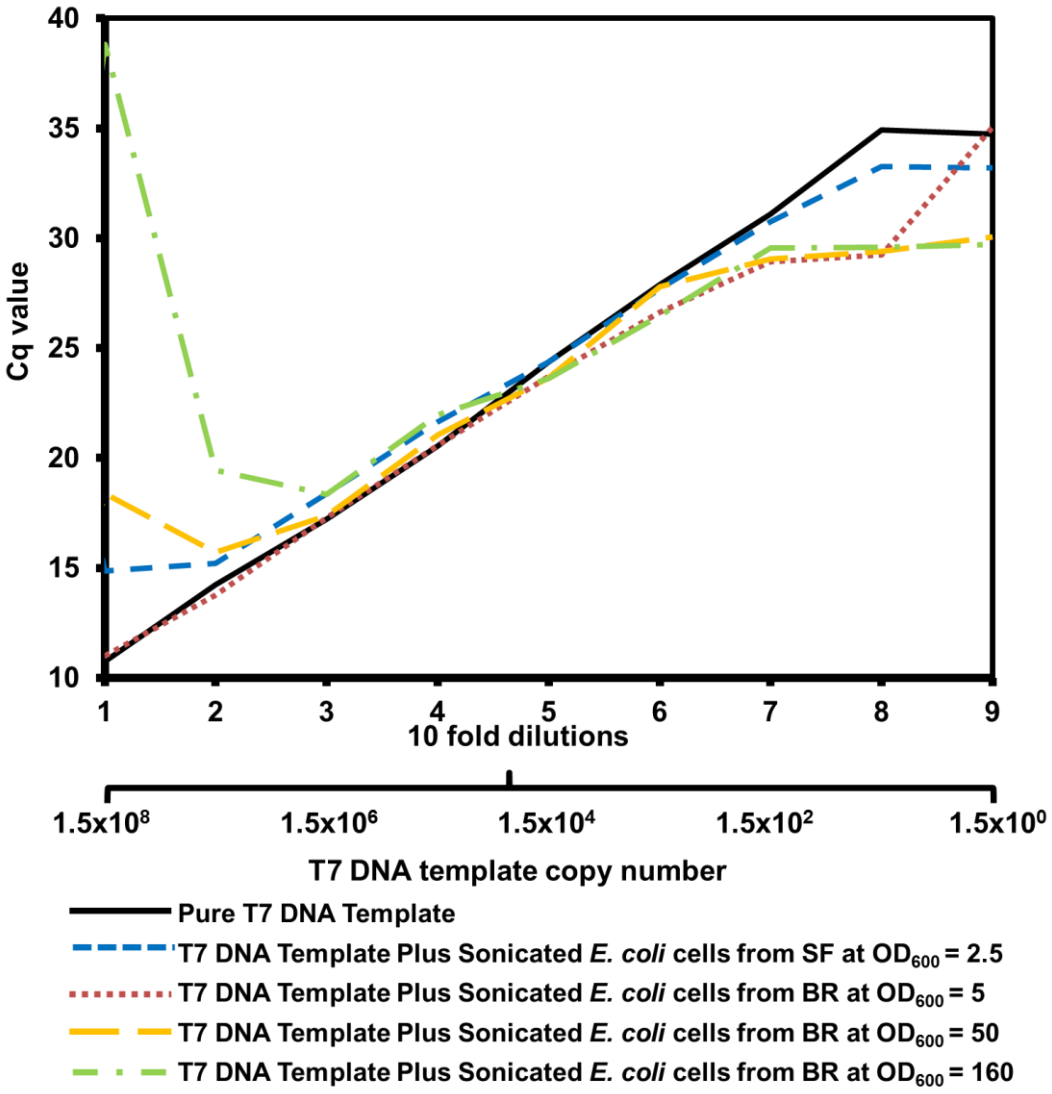

Figure 6-11 Influence of disrupted E. coli cells on amplification efficiency for a bacteriophage target sequence. Cq values were derived from real time PCR fluorescence data using a dilution series of template bacteriophage DNA either in pure water or the presence of disrupted cells derived from cultures of the indicated provenance and optical density.

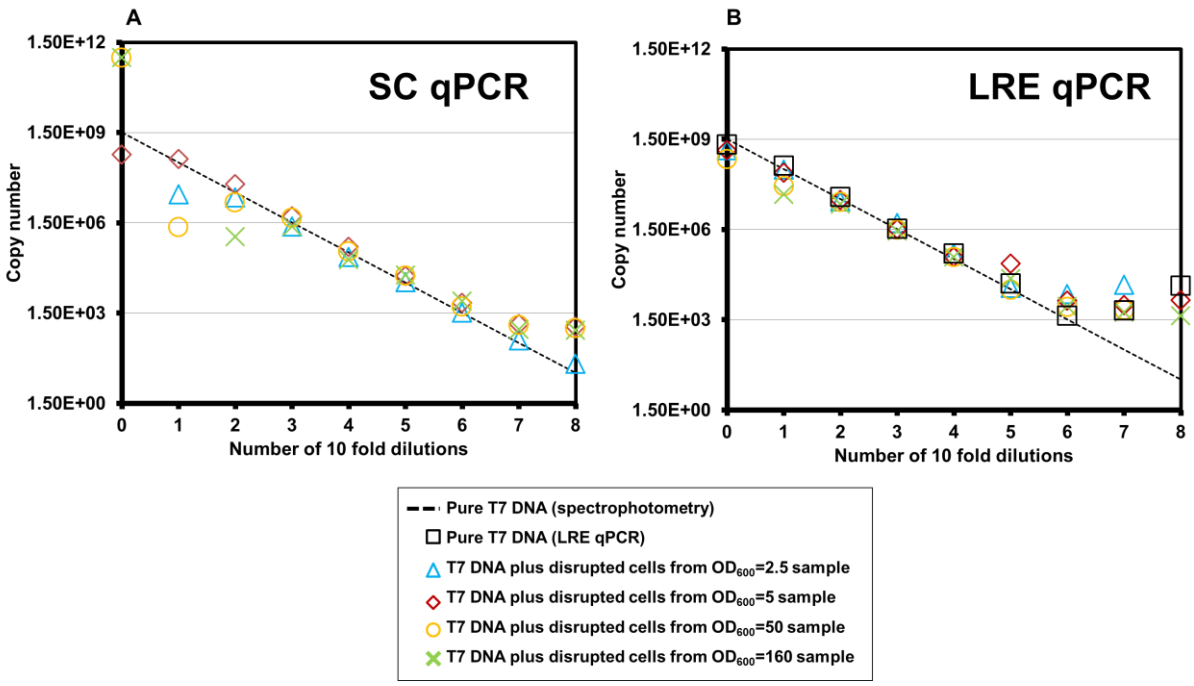

Figure 6-12 Comparison of SC qPCR and LRE-qPCR methods for absolute quantitation of a bacteriophage target sequence. Standard Curve (graph A) and LRE (graph B) methods were applied to real time PCR fluorescence data to quantify bacteriophage target sequence copy number either in purified DNA solution or plasmid DNA in the presence of disrupted cells derived from cultures of the indicated provenance and optical density.

6.3 Discussion

We were very keen to capture a clear data set that illustrates unambiguously the degree to which *E. coli* sample preparation procedures are required to capture accurate e-pPCR and qPCR data. If certain PCR-based assays require little or no sample preparation this has the potential to help accelerate assay turnaround, expanding the range of applications to at-line monitoring and scale down, automated mimics of large scale bioindustrial processes. Another motivating factor is to address the sometimes 'word of mouth' route by which molecular biology 'lore' can spread amongst researchers and replace this with rigorous data.

Detection of the *E. coli* genomic target by e-pPCR (Figure 6-3) was reduced tenfold by the presence of cell debris from both OD_{600}=2.5 shake flask and OD_{600}=160 bioreactor cultivation The overall level of amplicon production was marginally increased by the presence of bioreactor-derived disrupted cells (Figure 6-3 Bii) and decreased by the presence of disrupted cells from shake flasks (Figure 6-3 Aii). Detection of bacteriophage DNA by e-pPCR was reduced tenfold by the presence of OD_{600}=2.5 shake flask material and 100-fold by the OD_{600}=5 bioreactor material (Figure 6-9). Cellular material also reduced the level of bacteriophage amplicon production relative to naked DNA for both shake flask (Figure 6-9 Aii) and bioreactor-derived samples (Figure 6-9 Bii). These observations clearly support the utility of sample preparation for e-pPCR as a detection assay.

The presence of disrupted cells from shake flask cultivation did not impact the efficiency of amplification of a genomic target sequence (Figure 6-4 A) whereas disrupted cells from bioreactor cultivation did reduce amplification efficiency compared to that observed for purified DNA (Figure 6-4 B). For amplification of a bacteriophage target sequence, the presence of disrupted cells had little effect on amplification for samples that had undergone 3-7 tenfold dilutions. This holds true even for OD_{600}=160 starting material (Figure 6-11). Efficiency of amplification underlies many different statistical approaches to analysis of real time PCR data. As such, the finding that the presence of cellular material can in some cases have negligible effect on amplification

should encourage those researchers seeking to identify simple and rapid sample preparation procedures for qPCR.

Standardisation is a key pillar of both synthetic biology and the regulatory frameworks that constantly evolve to maximise safety in the biopharmaceutical industry. Ground-breaking work by Rutledge (2004)(Rutledge, 2004) suggested it is possible to have a universal standard for calibration and analysis of real time PCR data in the form of the LRE-qPCR method incorporating the Cal1-OCF reference reaction. We sought to test the robustness of LRE-qPCR by measuring the degree to which the presences of disrupted cells may compromise its accuracy. We also compare performance of LRE-qPCR to conventional 'standard curve' SC qPCR for absolute quantitation of a target sequence.

LRE-qPCR and SC qPCR were broadly equivalent in terms of quantifying genomic target DNA copy number in the presence of disrupted cells derived from shake flask or bioreactor cultivation (Figure 6-6). This was also borne out by statistical analysis (Figure 6-7). When applied to real time PCR for detection of bacteriophage DNA (Figure 6-12), LRE-qPCR appears to outperform SC qPCR in terms of agreement with spectrophotometry across all OD_{600} levels tested, except at very low template concentration (Figure 6-12).

Quantitation of a genomic target sequence by LRE-qPCR closely agreed with copy number estimates made using spectrophotometric data (Figure 6-8 A) for purified DNA and disrupted cell samples derived from shake flask cultivation. Undiluted samples from shake flasks contained in the order of 100-200ng DNA. By contrast, undiluted samples derived from shake flask cultivation contained 3.9µg-5µg of DNA and the resultant LRE-qPCR data (Figure 6-8 B) diverged from spectrophotometric data at the highest and lowest sample concentrations. We suggest DNA content has inhibited amplification at high concentrations. For both shake flask and bioreactor-derived material, LRE-qPCR measurements of copy number eventually plateau at low template concentration.

Our observations suggest LRE-qPCR can be used to quantify specific host cell DNA sequences in shake flask (Figure 6-6 A, Figure 6-8 A) and, with some dilution, bioreactor (Figure 6-6 B, Figure 6-8 B) process streams using only minimal sample preparation procedure, as described here. The same is also true for quantitation of bacteriophage DNA in process streams (Figure 6-12 B).

We have shown that the degree of sample preparation necessary for accuracy in PCR-based assays can vary both with the particular PCR method and with the provenance of the cells in the sample. We observed that sample preparation is certainly essential to maximise the sensitivity of e-pPCR - with unpurified samples showing 10-100 fold reduction in LOD compared to purified DNA samples. However, compromise of qPCR data caused by the presence of crude solutions of disrupted cells could be surmounted by straightforward sample dilution in most cases.

The accuracy profile of LRE-qPCR matches that of SC qPCR by the measures performed here. In light of this and previous validation of the properties of LRE-qPCR (Rutledge, 2004), we invite the synthetic biology community to use the Cal1-OCF standard and LRE procedure for absolute qPCR in anticipation that accumulation of data and experience could establish it as a valuable standard.

6.4 Conclusions

- Sample preparation remains a requirement for the e-pPCR analysis of genomic targets and targets separate to the host cell DNA, in this instance seeded plasmids representing bacteriophage infection.
- Disrupted cells from shake flask cultivation did not impact the efficiency of amplification of a genomic target sequence whereas disrupted cells from bioreactor cultivation reduced amplification efficiency, when compared to the same analysis on purified DNA .
- LRE-qPCR and SC qPCR were broadly equivalent in terms of quantifying genomic target DNA copy number in the presence of disrupted cells derived from shake flask or bioreactor cultivation.

- LRE-qPCR can be used to quantify genomic and plasmid DNA from a shake flask or bioreactor source (when diluted) with minimal sample preparation.
- Standardisation is a key pillar of Synthetic Biology and we suggest the CAL1 standard as a Synthetic Biology standard for qPCR.

Chapter 7

7 The CyCal curve - a Synthetic Biology standard for measuring host cell plasmids and genomes in *Escherichia coli*

7.1 Introduction

By monitoring genetic production factors in HCD fermentations, such as those contained within a genome or host cell plasmids, the bioprocess engineer can gain a better understanding of the processes underpinning biologics production and optimise fermentations to increase product titre (Ryu and Kim, 1993). Initiatives such as PAT have created further drivers for the production of timely data, specifically at-line or on-line data gathering and the adoption of synthetic biology in the field has created drivers for industrially robust standards (Gnoth et al., 2007). As shown in previous chapters, qPCR represents a solid platform on which to conduct such quantitative analyses of genetic information within fermentation.

As discussed in section 1.2.9, the "gold standard" SC qPCR method of absolute quantification has a number of limitations. These include increased practical complexity, a requirement for target specific analysis and the assumption of uniform amplification efficiencies across both standards and experimental samples. This assumption has often been shown to be false and makes the methodology unsuitable to any rapid sample processing that will leave inhibitors (such as cellular debris) within the reaction media (Miller et al., 1999). Furthermore standard curve construction itself is error prone, and small errors in the standard curve can translate into large errors in quantified copy number (Rutledge and Côté, 2003). These limitations serve to make absolute quantification impractical for studies examining even a small number of targets.

We now begin to address these issues by implementing knowledge gained from all previous studies and combining two alternative data analysis methods, in order to

accurately quantify multiple targets within sonicated process streams from a single non-specific standard curve. Previously we have applied various PCR data analysis methods quantify to genomic targets or seeded plasmid targets as proxies for contamination. We now attempt to refine our methodology to accurately quantify both genomic targets and host cell plasmids, which have been transformed and propagated intracellularly rather than seeded into the experimental sample.

7.2 Results

7.2.1 Theoretical development of the Cy0 Calibration (CyCal) Curve

The Cy0 method is applied to equalise efficiency between samples and standard and is described fully in section 1.6.2. Briefly, the Cy0 value is both a quantitative value and a reaction kinetic and is transformed into target quantity in the same manner as the Cq value. By transforming amplification curves via the Cy0 method the effect of inhibition on efficiency can be accounted for and of quantitative accuracy in samples where purity is compromised can be preserved.

The process of applying Cy0 led to the realisation that, as difference in efficiency necessitates the requirement of target-specific curves, a single non-specific curve could be applied to quantify samples where efficiency has been accounted for. As such we have adopted the CAL1 standard to produce the CyCal curve and logarithmically diluted the lambda phage DNA amplified by the CAL1 primers to generate the intercept and slope values. The CyCal curve can then be used to transform Cy0 values in the same manner that a standard curve is used to transform Cq values.

7.2.2 Comparison of CyCal quantification to Cq, Cy0 and LRE-qPCR on purified nucleic acid targets

As target specific standard curves remain the primary method with which to conduct absolute quantification, we have used both a target specific standard curve and the CyCal curve (constructed from logarithmically diluted lambda phage DNA) to quantify experimental targets diluted over a 5 fold dynamic range. These experimental samples

have been purified by phenol-chloroform extraction and should be relatively free from inhibitors, thus providing close agreement in copy number quantitation regardless of which data analysis method is used. From the amplification curves of these experimental samples, Cq and Cy0 values have been generated and both of these values have been transformed by the target specific curves and the CyCal curve. As a comparative measure, spectrophotometric measurements of the first purified sample were extrapolated into estimated copy number. Two experimental targets relevant to E. coli therapeutic biologic fermentation were quantified; a single copy gene on the E. coli chromosome, BirA, and a fragment antigen-binding (Fab) antibody gene contained within the pTODD A33 plasmid (Figure 7-1). The pTODD A33 plasmid is transformed into *E. coli* where it reaches a certain copy number and will allow for the production of Fab antibody fragments within the chassis upon induction. As the pTODD A33 plasmid has a ColE1 origin of replication, the expected copy number would be between 25-50 copies (Lutz et al, 1997).

Figure 7-1A shows the standard curve used for both the BirA and Fab targets and their application to determine the copy number of experimental samples. When constructing the target specific standard curves, spectrophotometry determined the first dilution of the BirA standard curve to contain 15ng of DNA and the first point of the Fab curve to contain 0.1ng of DNA. These were logarithmically diluted and linear regression of resultant Cq and log template mass gave the values for y-intercept and slope shown. When these standard curves are applied it can be seen that quantified BirA and Fab copy numbers show close agreement, regardless of quantification method used. It is notable that the Cy0 values show closer agreement to values for copy number determined by spectrophotometry than those from standard curves. This is likely due to the ability of Cy0 to remove the impact on efficiency caused by inhibitors, and small quantities of inhibitors will be co-purified during DNA extraction (Miller et al., 1999).

Figure 7-1B shows the CyCal curve and its application to quantifying the same Cy0 values used in Figure 7-1A. As the CyCal curve is not target specific it can be used to quantify Cy0 values generated from multiple targets. To construct the CyCal curve, an initial mass of 12ng of lambda DNA was serially diluted 7 times to give the values for

y-intercept and slope shown. LRE-qPCR was also used as it employs the same CAL1 reaction as a calibrator however uses it to generate an optical calibration factor (OCF) to link fluorescence to copy number. When profiled against CyCal, copy numbers derived from the CyCal curve are consistently higher than LRE-qPCR, however both show close agreement to the copy number estimated from spectrographic measurement.

We also directly compared copy numbers derived by transforming Cy0 values with a standard curve and Cy0 values with the CyCal curve, by calculating the percentage difference in copy number from the Cy0 value transformed by the standard curve and the Cy0 value transformed by the CyCal curve. It was noticed that marginal differences in copy number can be seen. The average difference in BirA copy number quantification across all dilutions is 13.5%. The average difference in Fab copy number estimation across all dilutions is 17.2% (Table 13). This indicates that the CyCal curve is able to provide a good agreement to standard curves when quantifying purified nucleic acid samples, despite it not being specific to either target.

Table 13 Percentage errors between copy number quantified using a target specific standard curve and copy number quantified using the CyCal curve

Dilution	Genomic	Host plasmid
1	18.53%	18.05%
2	16.16%	17.63%
3	13.58%	17.22%
4	10.96%	16.77%
5	8.29%	16.26%
Ave	**13.5%**	**17.2%**

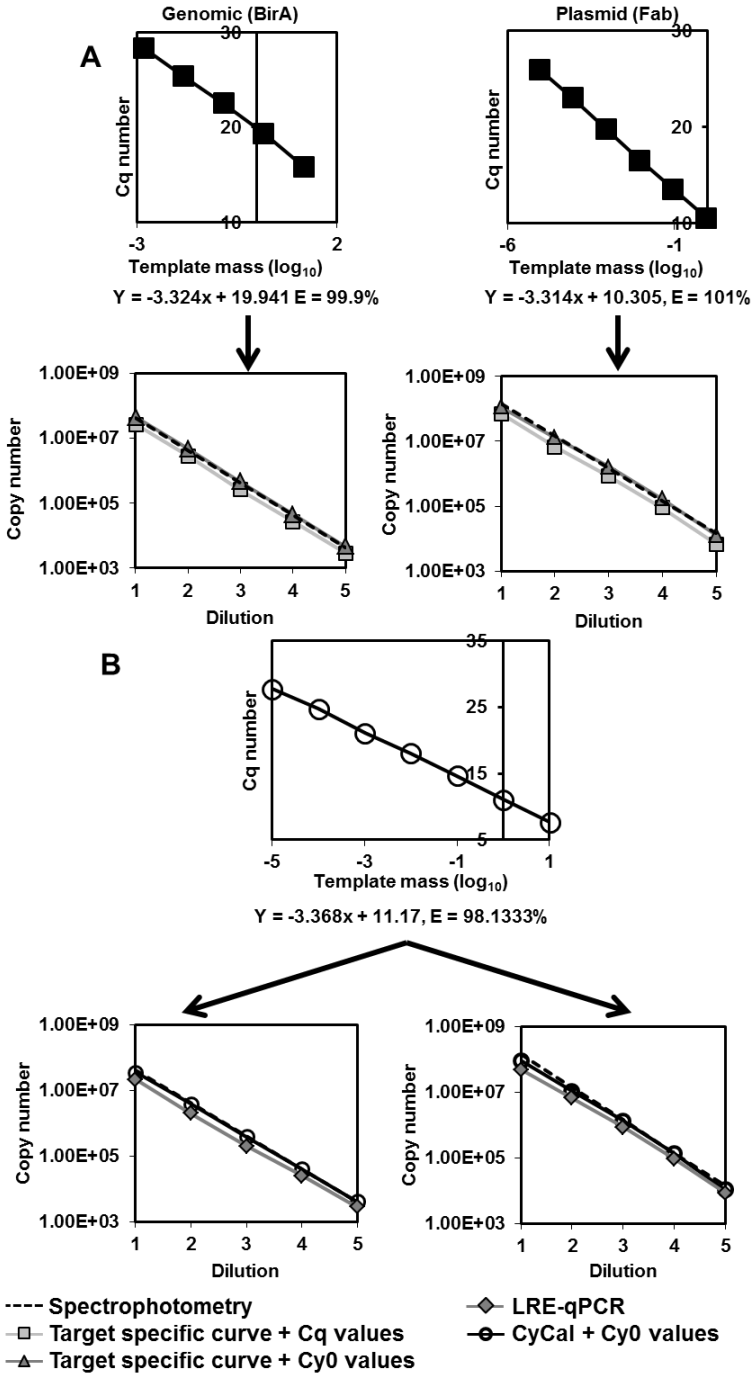

Figure 7-1 Comparison of target specific curves and the CyCal curve in the quantification of targets from purified E. coli gDNA and pTODD A33 pDNA. (A) Target specific curves were constructed from purified *E. coli* gDNA and pTODD A33 pDNA. Copy number derived from Cq and Cy0 transformation has been plotted against copy number estimated from spectrophotometry to profile accuracy of each method. (B) The CyCal curve was constructed using purified Bacteriophage λDNA. This was used to transform Cy0 values of both BirA and Fab' targets and the results plotted against extrapolated spectrophotometry data. LRE-qPCR, that used the CAL1 reaction as a calibrator was also used to quantify copy number directly from the amplification profiles of each target.

7.2.3 Development of a CyCal assay capable of monitoring PCN and host cell plasmids from crudely prepared samples

We next explore the CyCal curve's ability to monitor multiple targets within samples taken from shake flask fermentation, in order to track the number of copies of the production plasmid per cell. As extraction of DNA limits assay throughput and introduces error through loss of material, the assay has been performed on sonicated process streams using methodologies from previous studies. All methods of data analysis have been utilised to allow assessment of how this rapid sample preparation affects quantitative accuracy between the methods. We have calculated the PCN through the ratio of Fab to BirA. These experiments aim to quantify plasmids from the host cell and not plasmids that are seeded into the sample, as was done in previous experiments. Therefore plasmids can be supercoiled or conjugated to proteins, which could affect their measurement by qPCR.

Figure 7-4 shows that quantitation of sonicated process streams results in markedly different results depending on the data analysis method used. When quantifying a BirA target (Figure 7-2A) LRE-qPCR gives the lowest estimated copy number for BirA, indicating it is most affected by inhibition. There is very close agreement between CyCal-quantified and standard curve-quantified Cy0 values. We calculated average difference in error between Cy0 analysed with CyCal and Cy0 analysed with a target specific curve as 6.22%, demonstrating that the non-specific CyCal curve is able to produce copy numbers in close agreement to the target specific curve. When Cq values are transformed by target specific curves they fall between LRE-qPCR and CyCal/Cy0 values, which directly shows how the generation of Cy0 is able to account for loss of amplification efficiency in quantification.

The quantification of Fab by LRE-qPCR and Cq values transformed with a target specific curve produce similar data. This is in contrast to the large difference in copy number when the BirA target is analysed, suggesting impacted efficiency is not uniform between targets when quantifying by LRE-qPCR. Cy0 values transformed by target specific curves and CyCal curves again produce a copy number that is both close in agreement and higher than other methods used. The average difference in error between

Cy0 analysed with CyCal and Cy0 analysed with a target specific curve is 3.61%, further demonstrating that the non-specific CyCal curve has the capacity to produce quantitative data that is extremely close to a target specific curve.

The ratio of BirA copies and Fab copies can be used to provide an estimate of Fab copies per cell, or plasmid copy number. When this is performed it can be seen Cy0, Cq and CyCal produce comparable estimations of PCN. LRE-qPCR, as the effect of inhibition has been non-uniform between the targets, produces a markedly different quantity of PCN. The copy numbers determined from this method fall largely within the range of those provided in the literature for the ColE1 origin of replication (Lutz et al, 1997). The only exception is the PCN determined through LRE-qPCR, which is initially higher than the other methods and slightly higher than the literature reported copy number.

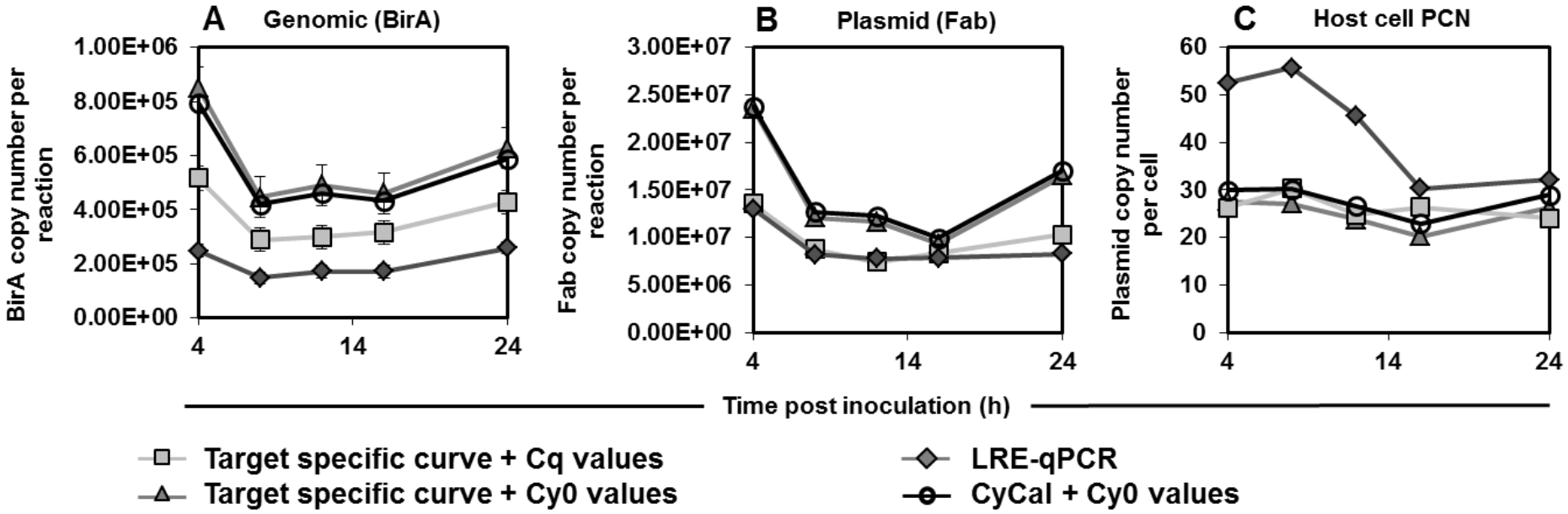

Figure 7-2 Quantification of BirA and Fab' targets from whole cell lysate of a shake flask fermentation, using a variety of qPCR data analysis techniques, and subsequent calculation of Fab' copies per cell. (A) Quantification of BirA gene copy number. (B) Quantification of Fab' fragment gene copy number. (C) Calculation of Fab' copies per cell, achieved by dividing total Fab' copy number by total BirA copy number.

7.2.4 Application of CyCal curve to quantification of genetic production factors in high cell density *E. coli* fermentation.

After showing that the CyCal curve can produce comparable data to target specific curves when analysing genomic DNA and host cell plasmids, we then used it to track the evolution of gDNA (BirA) and host cell plasmids (Fab) throughout HCD fermentation. In an industrial setting the monitoring of BirA evolution can be used to benchmark cell growth and monitoring of Fab evolution allows us to track how the quantity of the production gene changes over the course of the fermentation. Both of these can be combined to provide an estimate of how the PCN changes throughout fermentation. We have used flow cytometry as an alternative monitoring method to validate our data.

Monitoring was conducted post induction and at high optical densities of up to OD600=140 (Figure 7-3 A). Figure 7-3B shows BirA profiled against Fab'. Following an initial increase BirA is stable in the early stages of fermentation. In contrast Fab copies are seen to steadily increase, suggesting that *E. coli* is increasing its PCN at this stage (Nordstrom et al, 2006). As BirA copies begin to decrease later into the fermentation, Fab copies initially follow the same trend and then increase before resuming the same overall trend. PCN copies again largely falls into the range of 25-50 provided into the literature, with the exception of the later stages of fermentation when it exceeds this range.

BirA copies in the fermentation have been profiled against flow cytometry data in Figure 7-3C. Both should provide comparable quantities as, due to BirA being a single-copy gene, instances of it should correlate to a cell count. As can be seen the two methods share general agreement, however towards the end of fermentation the number of cells calculated using CyCal is slightly lower than those counted through flow cytometry. This provides validation for the approach of using single gene copy analysis as an indication of cell count in the fermentation. Despite the presence of cell debris being present and the curve being non-specific, a relatively accurate quantity of cells per mL can be elucidated.

PCN is also profiled against volumetric productivity quantified by HPLC (Figure 7-3D). These numbers should correspond as an increase in PCN should result in an increase in

volumetric productivity. It can be seen that both can be seen steadily rise throughout the fermentation, however productivity stabilises at approximately 88 hours whereas PCN continues to rise. This could be due to free plasmids in the media from cell lysis at the late stage of fermentation, which introduces inaccuracy. Accurate monitoring at this stage would require further investigation and could require a further stage of sample processing to remove cell-free plasmid

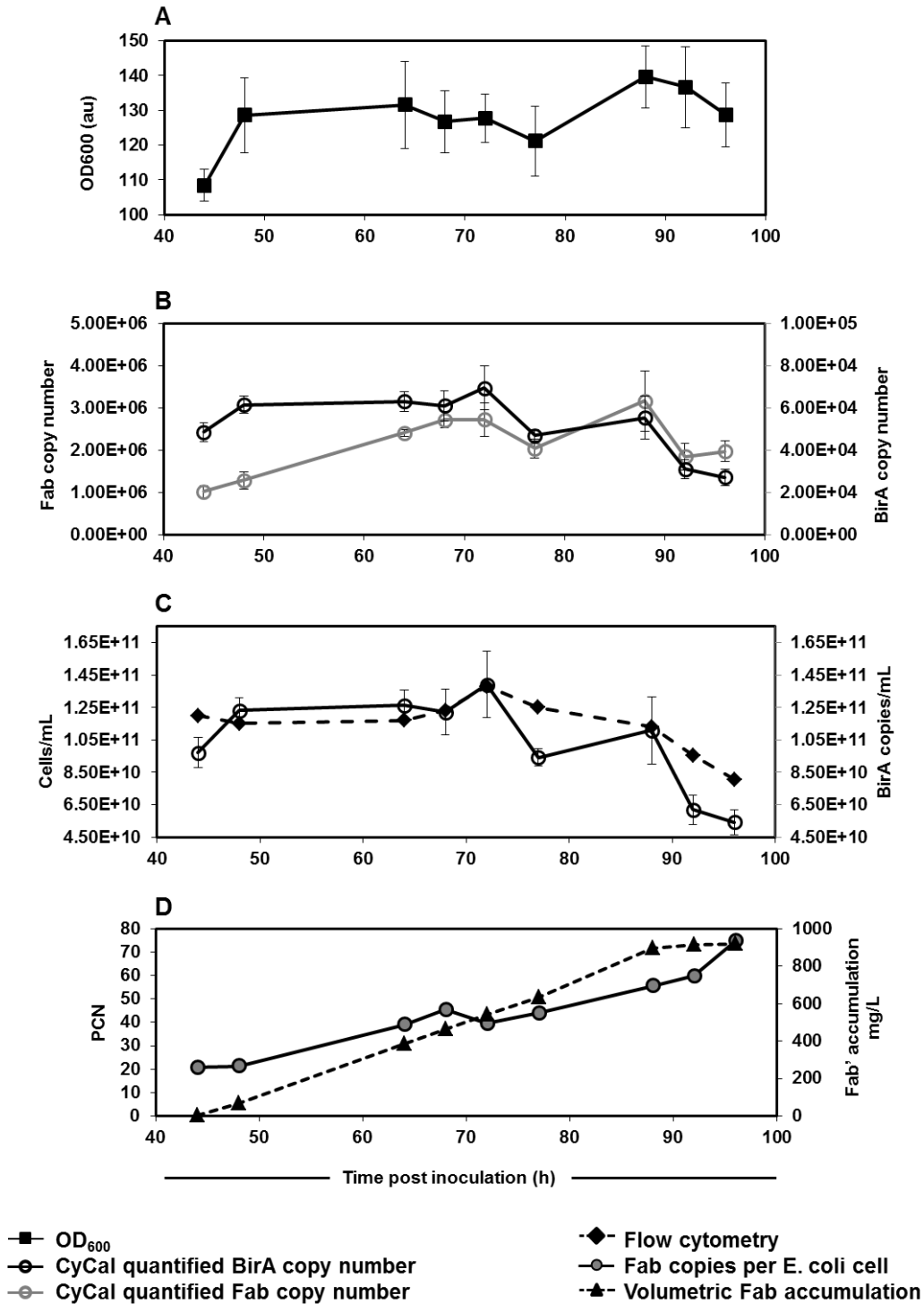

Figure 7-3 Application of CyCal curve to the monitoring of Fab' antibody fragment production in a high cell density E. coli fermentation. (A) Optical density. (B) Evolution of total BirA and Fab' copy numbers throughout the fermentation, as measured by application of CyCal curve to Cy0 values. (C) Comparison of BirA copy number from CyCal curve and Cy0 values and total viable cells as measured through flow cytometry. (D) Comparison of Fab' copies per cell, measured by dividing total Fab' copy number by total BirA copy number, against volumetric productivity data measured by HPLC analysis

7.2.5 Application of the CyCal curve to other bioprocessing targets

To further explore the potential of the CyCal curve as a bioprocess monitoring tool we have applied it to the quantification of several other purified and bioprocess relevant targets (Figure 7-4). These are a green fluorescence protein (GFP), often used as a reporter, contained within pJTDI and the bioprocess organisms used in previous studies; P. pastoris (Tkt genomic target) is targeted and CHO (GapDH genomic target).

Figure 7-4 shows that quantification of GFP copy number has close agreement between all methods and generally follows copy number expected by spectrophotometric measurement. The average difference between Cy0 values quantified from the standard curve and CyCal curve is 5.5%. Quantification of CHO and *P. pastoris* show a greater deviation in estimated copy number between methods. CyCal appears to slightly overestimate copy number when compared to spectrophotometric data for both of these targets. The average difference between standard curve quantified and CyCal quantified Cy0 values for P. pastoris and CHO is 30.3% and 56.3% respectively. This shows that more optimisation might be necessary for organisms with larger genomes.

Table 14 Percentage errors between copy number quantified using a target specific standard curve and copy number quantified using the CyCal curve

Dilution	Plasmid (GFP target)	Genomic *P. pastoris* (Tkt target)	Genomic CHO (GapDH)
1	32.49%	37.67%	28.65%
2	21.28%	33.69%	58.43%
3	8.16%	30.16%	58.17%
4	8.07%	26.67%	67.73%
5	26.55%	23.33%	68.46%
Ave	5.5%	30.3%	56.3%

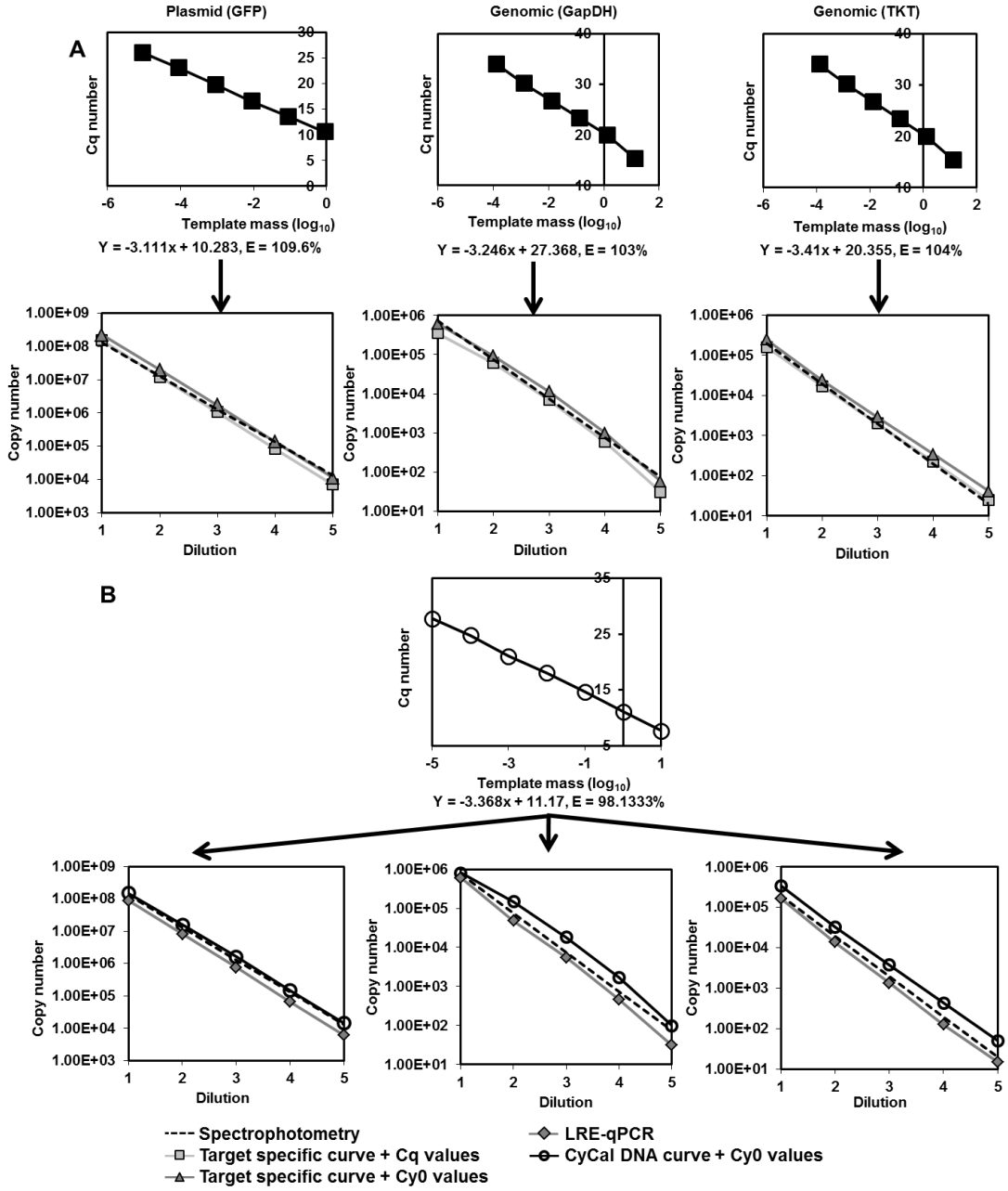

Figure 7-4Comparison of target specific curves and the CyCal curve in the quantification of targets from purified pJTDI pDNA and CHO and *P. pastoris* gDNA. (A) Target specific curves were constructed from logarithmically diluted purified DNA. Values for y-intercept and slope were applied to their specific targets to quantify copy number from Cq and Cy0 values and quantified copy numbers were plotted against values estimated from spectrophotometric measurements. (B) The CyCal curve was used to quantify all three targets from Cy0 values. LRE-qPCR, which also uses the same standard, was used to quantify copy number directly from the amplification curve. These methods were again plotted against values estimated from spectrophotometric measurements.

7.3 Discussion

Absolute quantification by qPCR remains the gold standard approach to analysing genetic targets. However the target specific curves required in current methodologies serve to introduce experimental and practical limitations, which have limited its widespread adoption in industrial settings.

Here we have achieved absolute quantification of multiple targets from a single CyCal curve constructed, which uses the CAL1 reaction components that we propose as a Synthetic Biology qPCR standard. We used the CyCal curve to simultaneously monitor two factors within high cell density *E. coli* fermentations, a genomic target and a host cell plasmid target. These were quantified from within sonicated process streams prepared using the methodology developed from previous studies. By monitoring the single copy genomic BirA gene the growth of *E. coli* can be tracked over the course of fermentation and the growth of the plasmid carrying the Fab productivity can also be tracked through its amplification. By determining the ratio of these targets to each other, the PCN, or number of host cell plasmids per genome, can be monitored.

PCN is an important metric in bioprocess monitoring as instability in the transfer of plasmids between dividing cells can lead to a population of cells lacking plasmids and therefore a loss of productivity (Friehs, 2004). The addition of antibiotics to the growth media can remedy this, however the added cost and impact on the final product quality at industrial-scale production makes their use impractical. Therefore it is desirable to maintain a high PCN, however an excessive plasmids within the cell can impact viability due to the imposition of metabolic burden (Skulj et al., 2008). Therefore plasmid vectors must be optimised and monitored to ensure their suitability to the application.

Currently, the primary means to monitor PCN are lengthy and complex procedures that use dangerous chemicals, such as caesium chloride centrifugation (Weisblum et al., 1979) and southern blot hybridisation (Olsson et al., 1993). Methods have been

developed that analyse a fluorescent reporter encoded within the plasmid. Lobner-Olesen (1999) developed a method that monitors PCN by inducing reporter expression for part of the cell cycle before blocking it before measuring fluorescence via flow cytometry. As plasmid copy number follows a stepwise rate of change, should show peaks of fluorescence distribution. The author however only witnessed smooth distributions and concluded that the method could at best detect two fold differences PCN, and was mainly suited to detecting large fluctuations in PCN (Loebner-Olesen, 1999). Wong et al. (2010) later adapted this, but a more suitable flow cytometry protocol was implemented and the reporter was induced for multiple cell cycles. Despite these changes, similar results to the earlier Loebner-Oleson method were obtained. Super-microscopy visualization of fluorescence has been attempted, but due to the requirement for a single bacterium to be fixed these have achieved limited success (Huang et al, 2009). Fluorescence in-situ hybridization approaches have also been attempted but there is currently no means of examining the fluorescence with the required resolution. This, coupled with the amount of non-specific targets present, mean the method is hampered by limitations (Zenklusen and Singer, 2010). As a result of these severe methodological limitations, assays using fluorescence reporters have failed to gain traction as suitable assays for determining PCN.

The Paulsson lab in 2012 discussed the limited means of measuring PCN and as such attempted to implement methods of determining plasmid loss rates. Although this is not a direct measure of the PCN it allows determination as to whether a plasmid is being partitioned unevenly during cell division. In order to do this they used both a probability-based method and analysis through microscopy, finding that plasmid loss rates are typically lower than reported. These however involves culturing the bacteria separately and is not as such a monitoring technique, and definitely not applicable to at-line analysis (Paulsson, 2012).

Due to these considerations, rapid determination of PCN is highly desirable feature of a bioprocess assay and the CyCal assay we demonstrate here circumvents these limitations. We did note that in the later stages of the fermentation, PCN did not

stabilize and exceeded literature reported copy numbers. This could be due to free plasmids being detected in the media and warrants further investigation.

It was noticed here that LRE-qPCR, found to have accuracy during our previous studies, was unable to accurately quantify the host cell plasmids and therefore unable to provide an accurate value for PCN. We speculate that this is due to the plasmid being propagated in the host cell rather than seeded into the experimental sample. From this it will be subject to supercoiling and association with host cell proteins, which could possibly interfere with analysis.

We next demonstrated the wider utility of the CyCal curve by successfully using it to quantify other bioprocess relevant targets, including the *P. pastoris* and CHO targets studied in previous chapters. Whilst these larger genomes will require further optimisation and validation to ensure accuracy, the initial copy numbers given are realistic and a promising start point for the assay.

The CyCal curve not only allows an increase of assay throughput for qPCR assays by removing the need for repeated preparation of purified standards, but also gives the proven ability to accurately quantify from samples where nucleic acid purification is not a requirement. This reduces sample preparation time and prevents unnecessary target loss. The standardisable nature of the CyCal curve makes the method industrially robust and allows it to conform to Synthetic Biology standards, and will hopefully facilitate the wider adoption of absolute quantification across the bioprocess community. Taking into account these factors, we propose the CyCal curve as a reproducible and standardisable calibration curve that could become the basis of industrially robust qPCR assays, allowing for the precise determination of absolute quantity of multiple bioprocess genetic targets within endogenous or exogenous host cell DNA.

With ultra-rapid qPCR technology becoming more prevalent in laboratories, we further propose that the CyCal curve can form the basis of an at-line assay compliant with PAT initiatives. The level of standardisation could be further increased by the

commercial supply of the CyCal standard from a centralised source, in line with other Synthetic Biology repositories such as those maintained by the BioBrick Foundation.

7.4 Conclusions

- The CyCal curve is a non-specific standard curve constructed from the CAL1 standard
- The CyCal curve was able to quantify a genomic E. coli target and a plasmid target with the same accuracy as other data analysis methods
- The CyCal curve was then used to quantify genomic *E. coli* and host cell plasmid targets in a crudely prepared shake flask sample and was found to have comparable-or-improved accuracy to other data analysis methods
- We developed an assay capable of monitoring PCN throughout HCD fermentation, which demonstrated accuracy when compared to other validation methods
- CyCal also shows efficacy when used to analyse other targets, however might need to be optimised for organisms with larger genomes
- We believe CyCal to be the basis of an industrially robust Synthetic Biology compliant qPCR monitoring standard

Chapter 8

8 General conclusions and recommendations for future work

The aims and objectives of this project, defined in section 1.7, have been largely achieved. Following the successful selection of targets and design of primers and testing of quantification methods on a purified target background (Chapter 3), rapid sample preparation by sonication was explored on three bioprocess relevant organisms, P. pastoris (Chapter 4), CHO (Chapter 5) and *E. coli* (Chapter 6) and common contaminants of CHO and *E. coli* fermentation. We also used this opportunity to explore alternative data analysis methods. From the knowledge gained through this we developed a standardised and high throughput assay, CyCal, to measure PCN in an *E. coli* fermentation (Chapter 7).

8.1 Implementation of rapid sample preparation

A common feature of PCR assays is the isolation of nucleic acid, which is done in order to separate out potential inhibitors that may be present in the growth media (Dineva et al., 2007). A number of strategies are available to achieve this, including manual extraction using a phenol-chloroform protocol and commercially available kits. These however serve to limit assay throughout time and have in some instances shown to introduce contaminants (Queipo-Ortuño et al., 2007). As such, a stated objective of this project was to measure the extent to which sample preparation was necessary across bioprocess relevant organisms and two simulated contaminating organisms, at either shake flask or bioreactor scale.

We concluded that both shake flask and bioreactor materials significantly impacted the ability of e-pPCR to detect both endogenous targets and simulated contaminants from both shake flask and bioreactor process streams. We consistently found a significant variation in amplicon production, as determined by densitometry. E-pPCR protocols are often used in detection of biological contaminants such as mycoplasmas (Kong et al, 2001; Sung et al., 2006; Dobrovolny and Bess, 2011), as accurate quantification is less

important than accurate detection and available protocols are able to distinguish between a range of species using one test. For these sensitivity is critical to avoid misleading results. Therefore our general conclusion is that e-pPCR detection of genomic and contaminant nucleic acid is not amenable to rapid sample preparation.

qPCR is widely employed to investigate the quantity of a nucleic acid target, in relative or absolute terms. The utility of these assays is very broad, and can range between analysis of genetic variation, quantification of host cells in process streams and detection of genetic leakage as part of quality control (Parodi et al., 2002; Voronin et al., 2009; Cong et al., 2013). Equipment platforms amplify the target and analyse the amplification in a single step and are very amenable to automation. Therefore it is perhaps absolute quantification by qPCR that is best served by a drastic reduction in sample preparation, as it occupies a much greater proportion of total assay time. We found overall that absolute quantification is achievable across a range of dilutions for all targets tested, with only a rapid sonication step totalling 5 minutes. We feel here that we have however defined the conditions at which absolute quantification is achievable through minimal sample preparation, for both shake flask and bioreactor scale, across three organism backgrounds and two contaminant target backgrounds.

8.2 Exploration of alternative qPCR data analysis methods

A plurality of methods of analysing qPCR data have been developed and two were explored here, as alternatives to the classical standard curve method of achieving absolute quantification by qPCR.

LRE-qPCR was implemented to complement the rapid sample preparation, as with the optical calibration it has the capacity to quantify nucleic acid target without the need for preparing a standard curve, thus further increasing assay throughput. We generally found when analysing a single target that LRE-qPCR performed in many instances with a comparable accuracy to SC-qPCR, although still failed to produce useable data at the most concentrated and dilute dilutions in many instances.

Upon attempting to analyse two targets in the same reaction, in order to quantify PCN, a small differential in the effect of inhibition and its impact quantification of each target meant that LRE-qPCR produced distorted results when quantifying PCN. As such a method of eliminating inhibition, Cy0, was implemented. This was adapted to use the Cal1 calibration reaction from LRE-qPCR to make the basis of a non-target specific reaction. We found that the implementation of Cy0 and its adaptation with the Cal1 standard could form the basis of a new, high throughput and standardised assay that meets the overarching aim of the project.

8.3 Introduction of Synthetic Biology standards

Synthetic Biology is striving to implement standards in order to ensure the diverse range components necessary for this nascent field are able to perform consistently and reproducibly, as the genetic parts are not fully predictable. Initiatives such as the Synthetic Biology Standards Consortium (Hayden, 2015) have been organised to address this and explore the implementation of such standards.

An objective of this project was to adapt this rationale into our investigations, in order to produce assays with standardised reaction components to ensure reproducibility. When implementing the LRE-qPCR data analysis method into our workflow we were pleased to see that the calibration steps used a standardised and non-target specifc calibration reaction, the Ca1 reaction, in order to generate an OCF. This was further expanded on by it's incorporation into the CyCal Curve.

Whilst PCR has standardised guidelines (Bustin et al., 2009), the components, targets and primers used feature no standardisation, especially when performing absolute quantification which relies on target specific standard curves (Rutledge et al., 2003). We therefore hoped to incorporate credible standards into PCR-based assays and feel we were successful in this endeavour in the production of the CyCal curve, which we promote as the first steps towards introducing such standards for widespread community use.

Assays such as the ones developed during this project serve to advance bioprocess monitoring and add to the toolbox of the synthetic biologist. Such rapid and simple to perform assays are highly desirable, as evidenced by initiatives such as PAT, which is creating drivers for at-line monitoring assays. If these assays were to be conducted on ultra-rapid qPCR platforms, such as the BJS Biotechnologies Xxpress, they could become the first at-line qPCR assays.

8.4 At-line and on-line monitoring for real-time data gathering in Bioprocessing

As the number and use of automated cell culture systems increases, the need for the real-time and high throughput analytics is increasing. The CyCal assay can be conceivably adapted into an assay capable of at-line monitoring, if used on ultra-rapid cycle qPCR platforms, allowing for close to real-time data gathering. This fits into the vision advanced by initiatives such as PAT, which promote at-line or on-line monitoring as a mean to study bioprocesses and advance their industrial capacity. The ability to gain results in (or close to) real-time is powerful when optimising fermentations and highly desirable to the bioprocess community.

With regard to other methods of gaining real-time information on genetic elements in fermentations through qPCR, advances in equipment are poised to achieve this. Specialised qPCR equipment is under development which allows the complete automation of qPCR assays within a closed workflow. Sampath et al. (2010) have developed such a system, which adapts an existing biosensor system (Ecker et al., 2006) to enable the fully automated detection of a number of adventitious contaminant microorganisms in bioprocess streams. Such devices are highly desirable as it allows a rapid response to such infection at industrial scale.

There are a number of technologies developing outside of qPCR that could provide at-line analysis of genetic elements in fermentations. Ultra-rapid sequencing combined with advances in bioinformatics mean the large scale sequencing of a sample combined with quantification of known sequences could become the standard method of quantification of genetic targets (Morozova and Marea, 2009) .

The development of on-line flow cytometry is another research avenue that could produce timely data on bioprocesses. Kuystermans and Avesh (2016) have developed a method using on-line flow cytometry that is capable of providing near real-time data on viable, early apoptotic, late apoptotic and necrotic cell populations within fermentation. As apoptosis is the primary source of cell death in fermentation, the gathering of this data is key improving control strategies for the production of biologics from cells.

A number of groups are developing whole cell biosensors for diagnostic or therapeutic applications, and it is highly conceivable that they could form a component of bioprocess monitoring. By reengineering cells to respond to environmental factors, most commonly through the expression of a sensor protein that couples to a ligand, a measurable response can be elicited. Examples of this come from Donaldson and Dattelbaum (2014), who took a transcriptional reporter assay and adapted into a whole-cell biosensor system. The result of this is that the presence of an analyte can be detected in the biosensor cell periplasm through an engineered periplasmic binding protein specific to the molecule. A number of protein interactions then follow that elicit a detectable fluorescent response. Interestingly, periplasmic binding proteins can be designed for a number of different analytes, creating a versatile biological monitoring system.

Soft sensors are another emerging strategy for on-line data gathering from bioprocess streams. A soft sensor is one or more hardware sensors combined with a mathematical model, which processes the signals generated by the sensors to deliver new data. They are commonly used to gain information on variables that cannot be otherwise obtained through on-line measurement (Mandenius and Gustavsson, 2015). Strategies have been devised using soft sensors, whereby sensors are incorporated at numerous downstream processing steps to continuously measure a variable throughout the process train. Through this, subsequent processing stages can be automatically configured based on the data gathered in the previous processing step (Velayudhan, 2014). The strategy of linking sensors capable of real-time data gathering to automatically gather data on critical process variables and adjust fermentation parameters accordingly is powerful in

its ability to influence product quality, reliability and consistency of production and positively impact production economic cost.

Exciting developments in monitoring technologies such as these bring the prospect of real-time gathering of bioprocess data closer to practicality. It is easy to envisage a future of fully automated bioprocessing, whereby the wealth of data contributed by associated monitoring technologies facilitates the rapid advancement of the field as a whole.

8.5 Future work

1) The CyCal curve can be further refined by using it to quantify different targets in an experimental setting and confirming accuracy with a wider range of validation techniques. The curve currently uses Cy0 to equalise efficiencies, however other alternative data analysis techniques could be explored.

2) It was noticed that PCN continues to escalate during the last stages of fermentation and we hypotheses that this could be due to the release of plasmids into the media from lysing cells. This hypothesis could be explored and rapid ways of isolating cells to negate this effect explored.

3) The CyCal assays are applicable to ultra-rapid platforms such as the BJS Biotechnologies Xxpress. An avenue of future research would be to gather data on platforms such as these and ensure accuracy is retained. SOP's for at-line qPCR assays could be developed.

4) For a standard to be adopted, it must be used by the wider community. Therefore the CyCal curve and any methodology could be released to other laboratories conducting qPCR to gather community data on its performance.

Appendix A

9 Publications

9.1 Templar, A., Woodhouse, S., Keshavarz-Moore, E., Nesbeth, D. N. (2016) Influence of Pichia pastoris cellular material on polymerase chain reaction performance as a synthetic biology standard for genome monitoring. *Journal of Microbiological Methods*, 127: 111-122.

9.2 Templar, A., Marsh, D., Nesbeth, D. N. (2016) A synthetic biology standard for Chinese Hamster Ovary cell genome monitoring and contaminant detection by polymerase chain reaction. *SpringerPlus*, Accepted for publication.

9.3 Templar A., Schofield, D., Nesbeth, D. N. (2016) Measuring E. coli and bacteriophage DNA in cell sonicates to evaluate the CAL1 reaction as a qPCR synthetic biology standard and end-point PCR. *Biomolecular Detection and Quantification.* Accepted for publication.

9.4 Schofield D., Templar, A., Newton, J., Nesbeth, D. N. (2016). Promoter engineering to optimize recombinant periplasmic Fab' fragment production in Escherichia coli. Biotechnology progress, 32: 840-847.

9.5 Templar A., Schofield D., Borg Y., Nesbeth D. N. (2016). Bacterial Cells as Engineered Chassis. In: Nesbeth D. N. ed. *Synthetic Biology Handbook*. Boca Raton: Taylor & Francis.

9.6 Templar A., Schofield D., Borg Y., Nesbeth D. N. (2016). Eukaryotae Synthetica: Synthetic Biology in Yeast and Mammalian Cells. In: Nesbeth D. N. ed. *Synthetic Biology Handbook*. Boca Raton: Taylor & Francis.

Appendix B – EngD Chapter

10 Commercial landscape of ultra-rapid qPCR

10.1 Commercial validation issues faced

BJS Biotechnology's sole business is the production of the Xxpress UF1 ultra-rapid thermocycler and its associated consumables. This is the first product that the company is entering into the market and was realised from the invention of a novel method of rapidly changing the temperature of the sample and reagents, which incorporates mechanisms that drive temperature change directly into the sample tray. There are a number of systems then built around this that must communicate and work in unison to allow for accurate data production and reporting. The two most crucial of these systems are an infra-red thermometer, which accurately detects this ultra-rapid change of temperature, and a system of cameras that records the change in fluorescence as the qPCR assay is carried out.

As PCR is an exponential reaction, small variances in reaction conditions can lead significant variances in results (Bar et al., 2003). Within the thermal cycler itself, thermal uniformity and accurate switching between temperatures are key to ensuring reaction conditions are optimal. There is a very low tolerance in accuracy for both of these factors before the validity of an assay is compromised through sub-optimal reaction conditions. A major challenge for BJS Biotechnology has been the design, implementation and validation of a system that is able to accurately and uniformly change temperature, whilst monitoring and reporting on temperature change, all at an extremely rapid rate. The thermometers used must take many readings a second and feed back to a system that must be capable of rapidly ramping temperatures up and down at a rate that is uniform across all samples on the plate. As this is the company's first product and the product is novel, the system has been designed and implemented from the ground up. This validation process is both extensive and ongoing, as it must be successfully reproduced, implemented and validated in every machine produced, to ensure accuracy across the whole Xxpress product line.

Regarding the equipment capable of monitoring fluorescence increase, the cameras used must be able to accurately distinguish between specific electromagnetic frequencies. This reflects the standard capacity of real-time PCR thermocyclers to "multiplex", whereby a number of fluorescent dyes are used that emit different frequencies to distinguish between separate genes in the same sample (Lehmann et al., 2008). The capability of the machine to accurately do this must be validated to ensure the machines commercial success.

Consumables also present a challenge for validation. There are a number of commercially available enzyme formulations (reagents) available to the PCR operator, each exhibiting slightly varying characteristics due to the polymerase used and proprietary contents of the formulation. Due to the extreme nature of ultra-rapid thermocycling, considerable pressure is exerted on the DNA polymerase and thus it is possible that some formulations will be incompatible with the Xxpress platform. Therefore a validation challenge is presented in testing which commercially available enzyme formulations are suitable.

These separate validation issues affect the success of the machine, as it must be able to reproduce results of commercial assays on machines already used in labs with equal or greater precision. A breakdown in the thermal ramping or fluorescence monitoring equipment will lead to inaccurate results. The use of unsuitable enzyme formulations can also introduce inaccuracy. As such the validation of these items is critical for BJS Biotechnology.

10.2 Research validation issues faced

Every ultra-rapid PCR assay developed over the course of this project must meet the current industry standards in terms of accuracy, sensitivity and throughput time. There must also demonstrably equal or hopefully superior performance to currently available assays that fulfil the same function. These assays could be PCR-based but might use a different technology. Assays developed for a clinical environment will be subject to an

especially rigorous validation process, as is required for any technology or product bought into the clinical environment.

The current scope of validation is to benchmark any at-line PCR assay developed for the Xxpress thermocycler against a similar offline PCR assay on an established thermocycler and a non-PCR assay used in industry. This will characterise performance against standard PCR techniques and equipment, as well as other assays currently used in industry. The data gathered from these experiments, if successful, can be used to argue the case for the adoption of at-line PCR assays to both industry and regulators.

In addition to validation, data generated by PCR assays must conform to a set of guidelines if it is to be reported on in the scientific literature. The Minimum Information for publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines were established by a panel of qPCR experts from academia and industry, due to the presence of incorrect and often un-reproducible qPCR data being found within the peer reviewed literature. Any data resulting from my project must have been generated in accordance to these guidelines in order for it to gain approval for publication.

References

Aaij, C., Borst, P., 1972. The gel electrophoresis of DNA. Biochimica et Biophysica Acta 269(2), pp. 192-200.

Abad, S., Kitz, K., Hörmann, A., Schreiner, U., Hartner, F.S., Glieder, A., 2010. Real-time PCR-based determination of gene copy numbers in Pichia pastoris. Biotechnology Journal. 5, pp. 413–420.

Agapakis, C.M., Silver, P.A., 2009. Synthetic Biology: exploring and exploiting genetic modularity through the design of novel biological networks. Molecular Biosyststems 5, pp. 704.

Annaluru, N., Muller, H., Mitchell, L.A., Ramalingam, S., Stracquadanio, G., Richardson, S.M., Dymond, J.S., Kuang, Z., Scheifele, L.Z., Cooper, E.M., Cai, Y., Zeller, K., Agmon, N., Han, J.S., Hadjithomas, M., Tullman, J., Caravelli, K., Cirelli, K., Guo, Z., London, V., Yeluru, A., Murugan, S., Kandavelou, K., Agier, N., Fischer, G., Yang, K., Martin, J.A., Bilgel, M., Bohutskyi, P., Boulier, K.M., Capaldo, B.J., Chang, J., Charoen, K., Choi, W.J., Deng, P., DiCarlo, J.E., Doong, J., Dunn, J., Feinberg, J.I., Fernandez, C., Floria, C.E., Gladowski, D., Hadidi, P., Ishizuka, I., Jabbari, J., Lau, C.Y.L., Lee, P.A., Li, S., Lin, D., Linder, M.E., Ling, J., Liu, J., Liu, J., London, M., Ma, H., Mao, J., McDade, J.E., McMillan, A., Moore, A.M., Oh, W.C., Ouyang, Y., Patel, R., Paul, M., Paulsen, L.C., Qiu, J., Rhee, A., Rubashkin, M.G., Soh, I.Y., Sotuyo, N.E., Srinivas, V., Suarez, A., Wong, A., Wong, R., Xie, W.R., Xu, Y., Yu, A.T., Koszul, R., Bader, J.S., Boeke, J.D., Chandrasegaran, S., 2014. Total Synthesis of a Functional Designer Eukaryotic Chromosome. Science 344, pp. 55–58.

Arabestani, M.R., Fazeli, H., Jedi Tehrani, M., Shokri, F. 2011. The Comparison of Microbial Culture and PCR Methods in Detection Of Cell Line To Mycoplasma. Journal of Isfahan Medical School. 28(121), pp. 1-8.

Arpino, J.A.J., Hancock, E.J., Anderson, J., Barahona, M., Stan, G.-B.V., Papachristodoulou, A., Polizzi, K., 2013. Tuning the dials of Synthetic Biology. Microbiology 159 pp. 1236–1253.

Baboo, J.Z., Galman, J.L., Lye, G.J., Ward, J.M., Hailes, H.C., Micheletti, M., 2012. An automated microscale platform for evaluation and optimization of oxidative bioconversion processes. Biotechnology Progress. 28, pp. 392–405.

Balasundaram, B., Nesbeth, D., Ward, J.M., Keshavarz-Moore, E., Bracewell, D.G., 2009. Step change in the efficiency of centrifugation through cell engineering: co-expression of Staphylococcal nuclease to reduce the viscosity of the bioprocess feedstock. Biotechnology and Bioengineering 104(1), pp. 134-42.

Bernstein, H.C., Carlson, R.P., 2012. Microbial Consortia Engineering for Cellular Factories: in vitro to in silico systems. Computational and Structural Biotechnology Journal. 3.

Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1, pp. 307–310.

Bravo, L.T.C., Procop, G.W., 2009. Recent advances in diagnostic microbiology. Seminars in Hematology. 46, pp. 248–258.

Burd, E.M., 2010. Validation of Laboratory-Developed Molecular Assays for Infectious Diseases. Clinical. Microbiology Review 23, pp. 550–576.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., 2009. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clinical. Chemistry. 55, pp. 611–622.

Chen, H., Horváth, C., 1995. Analytical Biotechnology: High-speed high-performance liquid chromatography of peptides and proteins. Journal of Chromatography. A 705, pp. 3–20. doi:10.1016/0021-9673(94)01254-C

Chervoneva, I., Li, Y., Iglewicz, B., Waldman, S., Hyslop, T., 2007. Relative quantification based on logistic models for individual polymerase chain reactions. Statistics in Medicine. 26, pp. 5596–5611.

Clark, W.A., Geary, D.H., 1974. The story of the American Type Culture Collection--its history and development (1899-1973). Advances in Applied Microbiology. 17, pp. 295– 309.

Clementschitsch, F., Bayer, K., 2006. Improvement of bioprocess monitoring: development of novel concepts. Microbial Cell Factories 5 pp. 19.

Clomburg, J.M., Gonzalez, R., 2010. Biofuel production in Escherichia coli: the role of metabolic engineering and Synthetic Biology. Applied Microbiology and Biotechnology. 86, pp. 419–434.

Cockerill, F.R., 2003. Application of rapid-cycle real-time polymerase chain reaction for diagnostic testing in the clinical microbiology laboratory. Archives of Pathology and Laboratory Medicine. 127, pp. 1112–1120.

Colin, V.L., Rodríguez, A., Cristóbal, H.A., 2011. The Role of Synthetic Biology in the Design of Microbial Cell Factories for Biofuel Production. Journal of Biomedicine and Biotechnology. 2011, pp. 1–9.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science 339, pp. 819–823.

Constantinou, A., Polizzi, K.M., 2013. Opportunities for bioprocess monitoring using FRET biosensors. Biochemical Society Transactions 41, pp. 1146–1151.

de Planell-Saguer, M., Rodicio, M.C., 2011. Analytical aspects of microRNA in diagnostics: A review. Analitica Chimica Acta 699, pp. 134–152.

De Schutter, K., Lin, Y.-C., Tiels, P., Van Hecke, A., Glinka, S., Weber-Lehmann, J., Rouzé, P., Van de Peer, Y., Callewaert, N., 2009. Genome sequence of the recombinant protein production host Pichia pastoris. Nature Biotechnology 27, pp. 561–566.

Demarest, S.J., Chen, G., Kimmel, B.E., Gustafson, D., Wu, J., Salbato, J., Poland, J., Elia, M., Tan, X., Wong, K., Short, J., Hansen, G., 2006. Engineering stability into Escherichia coli secreted Fabs leads to increased functional expression. Protein Engineering, Design & Selection 19, pp. 325–336.

De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouzé P, Van de Peer Y, Callewaert N. 2009. Genome sequence of the recombinant protein production host Pichia pastoris. Nature biotechnology, 27(6) pp. 561-566.

Dineva, M.A., Mahilum-Tapay, L., Lee, H., 2007. Sample preparation: a challenge in the development of point-of-care nucleic acid-based assays for resource-limited settings. Analyst 132, pp. 1193–1199.

Donaldson, T., Dattelbaum, J., 2014. Development of whole cell biosensors mediated by bacteria chemoreceptors. The FASEB Journal, 28(1) supplement 614.5.

Dymond, J.S., Richardson, S.M., Coombes, C.E., Babatz, T., Muller, H., Annaluru, N., Blake, W.J., Schwerzmann, J.W., Dai, J., Lindstrom, D.L., Boeke, A.C., Gottschling, D.E., Chandrasegaran, S., Bader, J.S., Boeke, J.D., 2011. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature 477, pp. 471–476.

Ecker, D.J., Drader, J., Gutierrez, J., Gutierrez, A., Hannis, J.C., Schink, A., Sampath, S., Blyn, L.B., Eshoo, M., W., Hall T.A., Tobarmosquera, M., Jiang, Y., Sannes-Lowery, K.A., Cummins, L. L., Libby, B., Walcott, J., Massire, C., Ranken, R., Manalili, S., Ivy, C., Melton, R., Levene, H., Harpin, V., Li, F., White, N., Pear, M.,. 2006. The Ibis T5000 Universal Biosensor: An Automated Platform for Pathogen Identification and Strain Typing. Journal of the Association for Laboratory Automation 11(6), pp. 341-351

Efiok, B.J.S., Eduok, E.E., International, A., 2000. Basic calculations for chemical and biological analysis. AOAC International.

Etemadzadeh, M. H., Arashkia, A., Roohvand, F., Norouzian, D., & Azadmanesh, K. 2015. Isolation, cloning, and expression of E. coli BirA gene for biotinylation applications. Advanced Biomedical Research, 4, 149.

Erlich, H.A., Gelfand, D., Sninsky, J.J., 1991. Recent advances in the polymerase chain reaction. Science 252(5013), pp. 1643-51.

FDA - Food and Drug Administration and Center for Biologics Evaluation and Research, Center for Biologics Evaluation and Research. 2010. Guidance for Industry: Characterization and Qualification of Cell Substrates and Other Biological Materials Used in the Production of Viral Vaccines for Infectious Disease Indications. US Department of Health and Human Services.

Ferre, F., 1992. Quantitative or semi-quantitative PCR: reality versus myth. Genome Research 2, pp. 1–9.

Freeman, W.M., Walker, S.J., Vrana, K.E., 1999. Quantitative RT-PCR: pitfalls and potential. BioTechniques 26, pp. 112–122, 124–125.

Friehs, K., 2004. Plasmid copy number and plasmid stability. Advances in Biochemical Engineering/Biotechnology 86, pp. 47–82.

Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E.A., Young, L., Qi, Z.-Q., Segall-Shapiro, T.H., Calvey, C.H., Parmar, P.P., Hutchison, C.A., Smith, H.O., Venter, J.C., 2010. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. Science 329, pp. 52–56.

Gil, M.E., Coetzer, T.L., 2004. Real-time quantitative PCR of telomere length. Molecular. Biotechnology. 27, pp. 169–172.

IMS Health, 2016. The Global Use of Medicines: Outlook through 2017. [Online] Available from: http://www.quotidianosanita.it/allegati/allegato1501906.pdf. [Accessed 03 June 2016]

Gnoth, S., Jenzsch, M., Simutis, R., Lubbert, A., 2007. Process Analytical Technology (PAT): Batch-to-batch reproducibility of fermentation processes by robust process operational design and control. Journal of Biotechnology 132, 2(31) pp. 180–186

Goll, R., Olsen, T., Cui, G., Florholmen, J., 2006. Evaluation of absolute quantitation by nonlinear regression in probe-based real-time PCR. BMC Bioinformatics 7, 107. doi:10.1186/1471-2105-7-107

Guescini, M., Sisti, D., Rocchi, M.B., Stocchi, L., Stocchi, V., 2008. A new real-time PCR method to overcome significant quantitative inaccuracy due to slight amplification inhibition. BMC Bioinformatics 9 pp. 326.

Guescini, M., Sisti, D., Rocchi, M.B.L., Panebianco, R., Tibollo, P., Stocchi, V., 2013. Accurate and Precise DNA Quantification in the Presence of Different Amplification Efficiencies Using an Improved Cy0 Method. PLoS ONE 8, e68481.

Guo, Y., Dong, J., Zhou, T., Auxillos, J., Li, T., Zhang, W., Wang, L., Shen, Y., Luo, Y., Zheng, Y., Lin, J., Chen, G.-Q., Wu, Q., Cai, Y., Dai, J., 2015. YeastFab: the design and construction of standard biological parts for metabolic engineering in Saccharomyces cerevisiae. Nucleic Acids Research. 43, e88.

Hacker, D.L., Baldi, L., Adam, M., Wurm, F.M., Flickinger, M.C., 2009. Transient Gene Expression in Mammalian Cells: Promises and Challenges for Medical Biotechnology, in: Encyclopedia of Industrial Biotechnology. John Wiley & Sons, Inc.

Hayden, E., 2015. Synthetic biology called to order: meeting launches effort to develop standards for fast-moving field. Nature 520, pp. 141–142

Hayduk, E.J., Lee, K.H., 2005. Cytochalasin D can improve heterologous protein productivity in adherent Chinese hamster ovary cells. Biotechnology and Bioengineering. 90, pp. 354–364.

Heinzle, E., 1992. High-Tech Analyses in Modern Biotechnology Present and potential applications of mass spectrometry for bioprocess research and control. Journal of Biotechnology. 25, pp. 81–114.

Heyland, J., Fu, J., Blank, L.M., Schmid, A., 2010. Quantitative physiology of Pichia pastoris during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. Biotechnology and Bioengineering. 107, pp. 357–368.

Higuchi, R., Dollinger, G., Walsh, P.S., Griffith, R., 1992. Simultaneous amplification and detection of specific DNA sequences. Nature Biotechnology. 10, pp. 413–417.

Higuchi, R., Fockler, C., Dollinger, G., Watson, R., 1993. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Nature Biotechnology 11, pp. 1026 - 1030

Huang, C.-J., Lin, H., Yang, X., 2012. Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. J. Ind. Microbiol. Biotechnol. 39, 383–399. doi:10.1007/s10295-011-1082-9

Huang B, Bates M, Zhuang X., 2009. Super resolution fluorescence microscopy. Annu Rev Biochem 78, pp. 993–1016.

Hyndman, D., Mitsuhashi, M., 2003. PCR Primer Design, in: Bartlett, J.S., Stirling, D. (Eds.), PCR Protocols, Methods in Molecular BiologyTM. Humana Press, pp. 81–88.

Inan, M., Fanders, S.A., Zhang, W., Hotez, P.J., Zhan, B., Meagher, M.M., 2007. Saturation of the secretory pathway by overexpression of a hookworm (Necator americanus) Protein (Na-ASP1). Methods in Molecular Biology. 389, 65–76.

Jayapal, K., Wlaschin, K., Hu, W., Yap, M., 2007. Recombinant protein therapeutics from CHO cells-20 years and counting. Chemical Engineering Progress 103, pp. 40–47.

Johnson, I.S., 1982. Authenticity and Purity of Human Insulin (recombinant DNA). Diabetes Care 5, pp. 4–12.

Kaiser, C., Pototzki, T., Luttmann, R., Ellert, A., 2008. Applications of PAT-Process Analytical Technology in Recombinant Protein Processes with Escherichia coli. Engineering in Life Sciences 8(2) pp. 132-138.

Kaltenboeck, B., Wang, C., 2005. Advances in real-time PCR: application to clinical laboratory diagnostics. Advances in Clinical Chemistry 40, pp. 219-259

Kantardjieff, A., Nissom, P.M., Chuah, S.H., Yusufi, F., Jacob, N.M., Mulukutla, B.C., Yap, M., Hu, W.-S., 2009. Developing genomic platforms for Chinese hamster ovary cells. Biotechnology. Advances 27(6) , pp. 1028–1035.

Katariina E. S. Tolvanen, P.E.P.K., 2008. Quantitative monitoring of a hydrogen-producing Clostridium butyricum strain from a continuous-flow, mixed culture bioreactor employing real-time PCR. International Journal of Hydrogen Energy 33, pp. 542–549.

Sikand, K., Singh, J., Ebron, J. S., & Shukla, G. C. (2012). Housekeeping Gene Selection Advisory: Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) and β-Actin Are Targets of miR-644a. PLoS ONE, 7(10), e47510.

Kelly, J.R., Rubin, A.J., Davis, J.H., Ajo-Franklin, C.M., Cumbers, J., Czar, M.J., Mora, K. de, Glieberman, A.L., Monie, D.D., Endy, D., 2009. Measuring the activity of BioBrick promoters using an in vivo reference standard. Journal of Biological Engineering 3, pp. 4.

Kim, J.Y., Kim, Y.-G., Lee, G.M., 2011. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Applied Microbiology and Biotechnology. 93, pp. 917–930.

Kitney, R., Freemont, P., 2012. Synthetic Biology – the state of play. FEBS Letters. 586, pp. 2029–2036.

Kiviharju, K., Salonen, K., Moilanen, U., Eerikäinen, T., 2008. Biomass measurement online: the performance of in situ measurements and software sensors. Journal of Industrial Microbiology and Biotechnology. 35, pp. 657–665.

Kong, F., James, G., Gordon, S., Zelynski, A., Gilbert, G.L., 2001. Species-Specific PCR for Identification of Common Contaminant Mollicutes in Cell Culture. Applied and Environmental Microbiology 67, pp. 3195–3200

Kononenko, A.V., Lee, N.C.O., Liskovykh, M., Masumoto, H., Earnshaw, W.C., Larionov, V., Kouprina, N., 2015. Generation of a conditionally self-eliminating HAC gene delivery vector through incorporation of a tTAVP64 expression cassette. Nucleic Acids Research 43(9).

Kristensen, T., Vestergaard, H., Møller, M.B., 2011. Improved Detection of the KIT D816V Mutation in Patients with Systemic Mastocytosis Using a Quantitative and Highly Sensitive Real-Time qPCR Assay. Journal of Molecular Diagnostics 13, pp. 180–188.

Kubista, M., Andrade, J.M., Bengtsson, M., Forootan, A., Jonák, J., Lind, K., Sindelka, R., Sjöback, R., Sjögreen, B., Strömbom, L., Ståhlberg, A., Zoric, N., 2006. The real-time polymerase chain reaction. Molecular Aspects of Medicine 27, pp. 95–125.

Kuystermans, D., Avesh, M. & Al-Rubeai, M. 2016. Online flow cytometry for monitoring apoptosis in mammalian cell cultures as an application for process analytical technology. Cytotechnology 68, pp. 399

Lajoie, M.J., Rovner, A.J., Goodman, D.B., Aerni, H.-R., Haimovich, A.D., Kuznetsov, G., Mercer, J.A., Wang, H.H., Carr, P.A., Mosberg, J.A., Rohland, N., Schultz, P.G., Jacobson, J.M., Rinehart, J., Church, G.M., Isaacs, F.J., 2013. Genomically Recoded Organisms Expand Biological Functions. Science 342, pp. 357–360.

Lau, B., Malkus, P., Paulsson, J., 2013. New quantitative methods for measuring plasmid loss rates reveal unexpected stability. Plasmid 70(3), pp. 353-361.

Lee, C., Kim, J., Shin, S.G., Hwang, S., 2006. Absolute and relative QPCR quantification of plasmid copy number in Escherichia coli. Journal of Biotechnology 123(3) pp. 273-280.

Lin Cereghino, G.P., Lin Cereghino, J., Jay Sunga, A., Johnson, M.A., Lim, M., Gleeson, M.A.G., Cregg, J.M., 2001. New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of Pichia pastoris. Gene 263, pp. 159–169.

Lincoln, C.K., Gabridge, M.G., 1998. Cell culture contamination: sources, consequences, prevention, and elimination. Methods in Cell Biology. 57, pp. 49–65.

Liu, W., Saint, D.A., 2002. Validation of a quantitative method for real time PCR kinetics. Biochemical and Biophysical Research Communications 294(2), pp. 347-353.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods 25, pp. 402–408.

Løbner-Olesen A.,1999. Distribution of minichromosomes in individual Escherichia coli cells: implications for replication control. EMBO J 18(6), pp. 1712.

Lopes, A. g., Keshavarz-Moore, E., 2012. Prediction and verification of centrifugal dewatering of P. pastoris fermentation cultures using an ultra scale-down approach. Biotechnology and Bioengineering. 109, pp. 2039–2047.

Lopes, A.G., Khan, N., Liddell, J., Keshavarz-Moore, E., 2012. An ultra scale-down approach to assess the impact of the choice of recombinant P. pastoris strain on dewatering performance in centrifuges. Biotechnology. Progress. 28, pp. 1029–1036.

Lutz R, Bujard H. 1997. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res. Mar 1525(6):1203-10

Macauley-Patrick, S., Fazenda, M.L., McNeil, B., Harvey, L.M., 2005. Heterologous protein production using the Pichia pastoris expression system. Yeast 22(4), pp. 249-70.

Mackay, I.M., 2004. Real-time PCR in the microbiology laboratory. European Journal of Clinical Microbiology & Infectious Diseases 10, pp. 190–212.

Malorny, B., Tassios, P.T., Rådström, P., Cook, N., Wagner, M., Hoorfar, J., 2003. Standardization of diagnostic PCR for the detection of foodborne pathogens. International Journal of Food Microbiology 83, pp. 39–48.

Mandenius CF., Gustavsson, R. 2015. Mini-review: soft sensors as means for PAT in the manufacture of bio-therapeutics. J Chem Technol Biotechnol, 90, pp. 215-227.

Mapelli, V., Olsson, L., Nielsen, J., 2008. Metabolic footprinting in microbiology: methods and applications in functional genomics and biotechnology. Trends in Biotechnology. 26, pp. 490–497.

Meijerink, J., Mandigers, C., van de Locht, L., Tönnissen, E., Goodsaid, F., Raemaekers, J., 2001. A novel method to compensate for different amplification efficiencies between patient DNA samples in quantitative real-time PCR. Journal of Molecular Diagnostics 3, pp. 55–61.

Miller, D.N., Bryant, J.E., Madsen, E.L., Ghiorse, W.C., 1999. Evaluation and Optimization of DNA Extraction and Purification Procedures for Soil and Sediment Samples. Applied Environmental. Microbiology. 65, pp. 4715–4724.

Miller, W.M., Blanch, H.W., Wilke, C.R., 1988. A kinetic analysis of hybridoma growth and metabolism in batch and continuous suspension culture: Effect of nutrient concentration, dilution rate, and pH. Biotechnology and Bioengineering. 32,pp. 947–965.

Morozova, O., Hirst, M., Marra, M.A., 2009. Applications of New Sequencing Technologies for Transcriptome Analysis. Annual Review of Genomics and Human Genetics 10, pp. 135–151.

Müller, K.M., Arndt, K.M., 2012. Standardization in Synthetic Biology. Methods in Molecular Biology 813, pp. 23–43.

Nesbeth, D.N., Perez-Pardo, M.-A., Ali, S., Ward, J., Keshavarz-Moore, E., 2012. Growth and productivity impacts of periplasmic nuclease expression in an Escherichia coli Fab' fragment production strain. Biotechnology and Bioengineering 109, pp. 517–527.

Nikfarjam L, Farzaneh P. 2012. Prevention and Detection of Mycoplasma Contamination in Cell Culture. Cell Journal;13(4), pp. 203-212.

Olsson, T., Ekwall, K., Ruusala, T., 1993. The silent P mating type locus in fission yeast contains two autonomously replicating sequences. Nucleic Acids Research. 21, pp. 855–861.

Nordström, K., & Dasgupta, S. 2006. Copy-number control of the Escherichia coli chromosome: a plasmidologist's view. EMBO Reports, 7(5), 484–489.

Ozturk, S.S., Palsson, B.Ø., 1990. Effects of Dissolved Oxygen on Hybridoma Cell Growth, Metabolism, and Antibody Production Kinetics in Continuous Culture. Biotechnology. Progress. 6, pp. 437–446.

Parodi, B., Aresu, O., Bini, D., Lorenzini, R., Schena, F., Visconti, P., Cesaro, M., Ferrera, D., Andreotti, V., Ruzzon, T., 2002. Species identification and confirmation of human and animal cell lines: a PCR-based method. Biotechniques.32(2), pp. 432-4, 436, 438-40.

Pavlov, A.R., Pavlova, N.V., Kozyavkin, S.A., Slesarev, A.I., 2004. Recent developments in the optimization of thermostable DNA polymerases for efficient applications. Trends in Biotechnology. 22(5) pp. 253-60.

Dobrovolny P. L., Bess, D., 2011. Optimised PCR detection of Mycoplasma. J Vis Ex (52), pp. 3057.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Research. ;29(9) pp. e45.

Pfaffl, M.W., Georgieva, T.M., Georgiev, I.P., Ontsouka, E., Hageleit, M., Blum, J.W., 2002. Real-time RT-PCR quantification of insulin-like growth factor (IGF)-1, IGF-1 receptor, IGF-2, IGF-2 receptor, insulin receptor, growth hormone receptor, IGF-153 binding proteins 1, 2 and 3 in the bovine species. Domestic Animal Endocrinology. 22, pp. 91–102.

Invitrogen: Pichia Fermentation Process Guidelines. Invitrogen 2002, Version B 053002.

Pioch, D., Jürgen, B., Evers, S., Maurer, K.-H., Hecker, M., Schweder, T., 2007. At-line Monitoring of Bioprocess-Relevant Marker Genes. Engineering in Life Sciences 7, pp. 373–379.

Queipo-Ortuño, M.I., Tena, F., Colmenero, J.D., Morata, P., 2007. Comparison of seven commercial DNA extraction kits for the recovery of Brucella DNA from spiked human serum samples using real-time PCR. European Journal of Clinical Microbiology & Infectious Diseases. 27, pp. 109–114.

Rasmussen, R., 2001. Quantification on the LightCycler, in: Meuer, P.D. med S., Wittwer, P.D.C., Nakagawara, D.K.-I. (Eds.), Rapid Cycle Real-Time PCR. Springer Berlin Heidelberg, pp. 21–34.

Riesenberg, D., Guthke, R., 1999. High-cell-density cultivation of microorganisms. Applied Microbiology and Biotechnology 51(4), pp. 422-430

Rios, M., 2012. A decade of microbial fermentation. BioProcess International 10, pp. 14-17.

Rutledge, R. G., Stewart, D., 2008. A kinetic-based sigmoidal model for the polymerase chain reaction and its application to high-capacity absolute quantitative real-time PCR. BMC Biotechnology. 8, pp. 47.

Ruijter, J.M., Pfaffl, M.W., Zhao, S., Spiess, A.N., Boggy, G., Blom, J., Rutledge, R.G., Sisti, D., Lievens, A., De Preter, K., Derveaux, S., Hellemans, J., Vandesompele, J., 2013. Evaluation of qPCR curve analysis methods for reliable biomarker discovery: Bias, resolution, precision, and implications. Methods 59, pp. 32–46.

Ruijter, J.M., Ramakers, C., Hoogaars, W.M.H., Karlen, Y., Bakker, O., Van Den Hoff, M.J.B., Moorman, A.F.M., 2009. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Research. 37(6) pp. e45

Ruth, C., Glieder, A., 2010. Perspectives on Synthetic Promoters for Biocatalysis and Biotransformation. ChemBioChem 11, pp. 761–765.

Rutledge, R.G., Côté, C., 2003. Mathematics of quantitative kinetic PCR and the application of standard curves. Nucleic Acids Research. 31, pp. e93–e93.

Rutledge, R.G., 2004. Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications. Nucleic Acids Research. 32, pp. e178–e178.

Rutledge, R.G., Stewart, D., 2008. Critical evaluation of methods used to determine amplification efficiency refutes the exponential character of real-time PCR. BMC Molecular. Biology. 9, pp. 96.

Rutledge, R.G., Stewart, D., 2010. Assessing the Performance Capabilities of LRE-Based Assays for Absolute Quantitative Real-Time PCR. PLoS ONE 5, e9731.

Rutledge, R.G., 2011. A Java Program for LRE-Based Real-Time qPCR that Enables Large-Scale Absolute Quantification. PLoS ONE 6, e17636.

Ryu, D.D.Y., Kim, J.Y., 1993. Bioprocess Kinetics and Optimization of Recombinant Fermentation: Genetic and Engineering Approaches, in: Yoshida, P.T., Tanner, P.D.R.D. (Eds.), Bioproducts and Bioprocesses 2. Springer Berlin Heidelberg, pp. 103–111.

Sandén, A.M., Prytz, I., Tubulekas, I., Förberg, C., Le, H., Hektor, A., Neubauer, P., Pragai, Z., Harwood, C., Ward, A., Picon, A., De Mattos, J.T., Postma, P., Farewell, A., Nyström, T., Reeh, S., Pedersen, S., Larsson, G., 2003. Limiting factors in Escherichia coli fed-batch production of recombinant proteins. Biotechnology and Bioengineering. 81, pp. 158–166.

Sampath R, Blyn LB, Ecker DJ., 2010. Rapid molecular assays for microbial contaminant monitoring in the bioprocess industry. PDA J Pharm Sci Technol. 2010;64(5), pp. 458–64.

Schrader, C., Schielke, A., Ellerbroek, L., Johne, R., 2012. PCR inhibitors – occurrence, properties and removal. Journal of Applied Microbiology. 113, pp. 1014–1026.

Shacter, E., Williams, J.A., Lim, M., Levine, R.L., 1994. Differential susceptibility of plasma proteins to oxidative modification: Examination by western blot immunoassay. Free Radical Biology & Medicine. 17, pp. 429–437.

Shendure, J., Ji, H., 2008. Next-generation DNA sequencing. Nature. Biotechnology. 26, pp. 1135–1145.

Shokere, L.A., Holden, M.J., Ronald Jenkins, G., 2009. Comparison of fluorometric and spectrophotometric DNA quantification for real-time quantitative PCR of degraded DNA. Food Control 20, pp. 391–401.

Singer, V.L., Lawlor, T.E., Yue, S., 1999. Comparison of SYBR® Green I nucleic acid gel stain mutagenicity and ethidium bromide mutagenicity in the Salmonella/mammalian microsome reverse mutation assay (Ames test). Mutation Research 439(1), pp. 37-47.

Skulj, M., Okrslar, V., Jalen, S., Jevsevar, S., Slanc, P., Strukelj, B., Menart, V., 2008. Improved determination of plasmid copy number using quantitative real-time PCR for monitoring fermentation processes. Microbial Cell Factories 7, pp. 6.

Soley, A., Lecina, M., Gámez, X., Cairó, J.J., Riu, P., Rosell, X., Bragós, R., Gòdia, F., 2005. On-line monitoring of yeast cell growth by impedance spectroscopy. Journal of Biotechnology. 118, pp. 398–405.

Sreekrishna, K., Potenz, R.H.B., Cruze, J.A., McCombie, W.R., Parker, K.A., Nelles, L., Mazzaferro, P.K., Holden, K.A., Harrison, R.G., Wood, P.J., Phelps, D.A., Hubbard, C.E., Fuke, M., 1988. High level expression of heterologous proteins in methylotrophic yeast Pichia pastoris. Journal of Basic Microbiology. 28, pp. 265–278.

Sturino, J.M., Klaenhammer, T.R., 2006. Engineered bacteriophage-defence systems in bioprocessing. Nature Reviews Microbiology 4, pp. 395–404.

Sung H, Kang SH, Bae YJ, Hong JT, Chung YB, Lee CK, Song S. 2006. PCR-based detection of Mycoplasma species. J Microbiol; 44(1):, pp. 42-49.

Teixeira, A.P., Oliveira, R., Alves, P.M., Carrondo, M.J.T., 2009. Advances in on-line monitoring and control of mammalian cell cultures: Supporting the PAT initiative. Biotechnology Advances 27, pp. 726–732.

Tichopad, A., Dilger, M., Schwarz, G., Pfaffl, M.W., 2003. Standardized determination of real-time PCR efficiency from a single reaction set-up. Nucleic Acids Research 31(20), pp. e122.

Timenetsky, J., Santos, L.M., Buzinhani, M., Mettifogo, E., 2006. Detection of multiple mycoplasma infection in cell cultures by PCR. Brazilian Journal Of Medical and Biological Research 39, pp. 907–914.

Tolner, B., Smith, L., Begent, R.H.J., Chester, K.A., 2006. Production of recombinant protein in Pichia pastoris by fermentation. Nature Protocols. 1, pp. 1006–1021.

Tustian, A.D., Salte, H., Willoughby, N.A., Hassan, I., Rose, M.H., Baganz, F., Hoare, M., Titchener-Hooker, N.J., 2007. Adapted ultra scale-down approach for predicting the centrifugal separation behavior of high cell density cultures. Biotechnology Progress. 23, pp. 1404–1410.

Velayudhan A. 2014. Overview of integrated models for bioprocess engineering. Curr Opin Chem Eng 6, pp. 83-89.

Vallone, P.M., Butler, J.M., 2004. AutoDimer: a screening tool for primer-dimer and hairpin structures. BioTechniques 37, pp. 226–231.

VanGuilder, H.D., Vrana, K.E., Freeman, W.M., 2008. Twenty-five years of quantitative PCR for gene expression analysis. BioTechniques 44, 619–626.

Vehar, G.A., Spellman, M.W., Keyt, B.A., Ferguson, C.K., Keck, R.G., Chloupek, R.C., Harris, R., Bennett, W.F., Builder, S.E., Hancock, W.S., 1986. Characterization Studies of Human Tissue-type Plasminogen Activator Produced by Recombinant DNA Technology. Cold Spring Harbor Symposia on Quantitative Biology 51, pp. 551–562.

Velez-Suberbie, M.L., Tarrant, R.D.R., Tait, A.S., Spencer, D.I.R., Bracewell, D.G., 2013. Impact of aeration strategy on CHO cell performance during antibody production. Biotechnology Progress 29, pp. 116–126.

Voronin, Y., Holte, S., Overbaugh, J., Emerman, M., 2009. Genetic Drift of HIV Populations in Culture. PLoS Genetics. 5.

Walsh, G., 2014. Biopharmaceutical benchmarks 2014. Nature Biotechnology 32, pp. 992–1000.

Walsh, G., Jefferis, R., 2006. Post-translational modifications in the context of therapeutic proteins. Nature Biotechnology. 24, pp. 1241–1252.

Wang, L., Brock, A., Herberich, B., Schultz, P.G., 2001. Expanding the genetic code of Escherichia coli. Science 292, pp. 498–500.

Wang, Y., Chu, J., Zhuang, Y., Wang, Y., Xia, J., Zhang, S., 2009. Industrial bioprocess control and optimization in the context of systems biotechnology. Biotechnology Advances 27, pp. 989–995.

Weber, W., Link, N., Fussenegger, M., 2006. A genetic redox sensor for mammalian cells. Metabolic Engineering 8, pp. 273–280.

Weisblum, B., Graham, M.Y., Gryczan, T., Dubnau, D., 1979. Plasmid copy number control: isolation and characterization of high-copy-number mutants of plasmid pE194. Journal of Bacteriology 137, pp. 635–643.

Wilhelm, J., Pingoud, A., 2003a. Real-time polymerase chain reaction. Chembiochem 4, pp. 1120–1128.

Wilson, K., 2001. Preparation of Genomic DNA from Bacteria, in: Current Protocols in Molecular Biology. John Wiley & Sons, Inc.

Wong Ng J, Chatenay D, Robert J, Piorer MG., 2010. Plasmid copy number noise in monoclonal populations of bacteria. Phys Rev E 81.

Xia, X.-X., Qian, Z.-G., Ki, C.S., Park, Y.H., Kaplan, D.L., Lee, S.Y., 2010. Native-sized recombinant spider silk protein produced in metabolically engineered Escherichia coli results in a strong fiber. Proceedings of the National Academy of Sciences107, pp. 14059–14063.

Zenklusen D, Singer RH., 2010. Analyzing mRNA expression using single mRNA resolution fluorescent in situ hybridization. Method enzymol 470, pp. 641–659.

Zhang, H., Wang, W., Quan, C., Fan, S., 2010. Engineering Considerations for Process Development in Mammalian Cell Cultivation. Current Pharmaceutical Biotechnology 11, pp. 103–112.

Zhang, M., Wang, H.Y., 1994. Hydrogen peroxide production using chemically treated Pichia pastoris cells. Enzyme and Microbial Technology. 16, pp. 10–17