UNIVERSITYOF BIRMINGHAM

Research at Birmingham

Use of a stress-minimisation paradigm in high cell density fed-batch Escherichia coli fermentations to optimise recombinant protein production.

Wyre, Christopher; Overton, Timothy

Document Version
Peer reviewed version

Link to publication on Research at Birmingham portal

1   **Use of a stress-minimisation paradigm in high cell density fed-batch *Escherichia coli* fermentations**

2   **to optimise recombinant protein production**

3   Chris Wyre and Tim W Overton*

4   Bioengineering, School of Chemical Engineering, and Institute of Microbiology & Infection, University

5   of Birmingham, Edgbaston, Birmingham B15 2TT, UK

6   *Corresponding author: t.w.overton@bham.ac.uk

7   Telephone: +44 (0) 121 414 5306

8   Fax: +44 (0) 121 414 5324

9   **ABSTRACT**

10   Production of recombinant proteins is an industrially important technique in the biopharmaceutical

11   sector. Many recombinant proteins are problematic to generate in a soluble form in bacteria as they

12   readily form insoluble inclusion bodies. Recombinant protein solubility can be enhanced by

13   minimising stress imposed on bacteria through decreasing growth temperature and the rate of

14   recombinant protein production. In this study, we determined whether these stress minimisation

15   techniques can be successfully applied to industrially-relevant high cell density *Escherichia coli*

16   fermentations generating a recombinant protein prone to forming inclusion bodies, CheY-GFP. Flow

17   cytometry was used as a routine technique to rapidly determine bacterial productivity and

18   physiology at the single cell level, enabling determination of culture heterogeneity. We show that

19   stress-minimisation can be applied to high cell density fermentations (up to a dry cell weight of > 70

20   g·L^{-1}) using semi-defined media and glucose or glycerol as carbon sources, and using early or late

21   induction of recombinant protein production, to produce high yields (up to 6 g·L^{-1}) of aggregation-

22   prone recombinant protein in a soluble form. These results clearly demonstrate that stress

23   minimisation is a viable option for the optimisation of high cell density industrial fermentations for

24 the production of high yields of difficult-to-produce recombinant proteins, and present a workflow

25 for the application of stress minimisation techniques in a variety of fermentation protocols.

26 **Keywords:** Green fluorescent protein; fed-batch fermentation; flow cytometry; inclusion bodies

27

INTRODUCTION

29 Production of recombinant proteins in bacterial hosts is an important part of the biopharmaceutical

30 industry. Although many new biopharmaceutical drugs (such as monoclonal antibodies) are made in

31 mammalian hosts, bacterial hosts are undergoing a resurgence in popularity not only for relatively

32 simple protein and peptide drugs (such as insulin) but also for more complex molecules such as

33 antibody fragments. Of the 58 biopharmaceutical drugs approved between 2006 and June 2010, 17

34 were produced in *E. coli* [29]. Efficient production of soluble recombinant proteins is also essential

35 for generation of proteins for structural studies that are the basis of the development of new drug

36 ligand molecules.

37 Two major routes are utilised for recombinant protein production in bacterial hosts. First, the

38 recombinant protein can be generated in the form of inclusion bodies (IBs); dense intracellular

39 particles comprising mainly misfolded protein but also containing some correctly folded, functional

40 protein [6]. Many recombinant proteins have a tendency to form IBs in bacterial hosts for a number

41 of reasons including differences in folding pathways and physicochemical conditions between

42 bacteria and eukaryotic cells [14]. Inclusion bodies are relatively simple to generate to high yields,

43 are easy to purify due to their density, and are a relatively pure source of recombinant protein [21].

44 However, to generate functional protein, IBs must be denatured (usually chemically) and then the

45 protein must be refolded; this refolding step varies in success and yield, such that some recombinant

46 proteins cannot be successfully refolded.

2

47   Alternatively, the recombinant protein may be synthesised in a correctly-folded, soluble form, and

48   purified in this functional form [22]. This is especially desirable for recombinant proteins with

49   multiple isoforms and for structural determination studies. However, there are many potential

50   problems with this approach, and some recombinant proteins cannot be readily synthesised in

51   bacteria in a soluble form. Often, attempts to produce soluble recombinant protein lead to IB

52   formation, decreases in host cell viability, loss of recombinant protein-encoding plasmids and

53   overgrowth of non-producing, plasmid-free bacteria, all of which can lead to low biomass and

54   recombinant protein yields. Indeed, many recombinant proteins that are difficult to produce are

55   referred to as 'toxic' proteins, due to the apparently toxic effects on host bacteria [17]. Nonetheless,

56   soluble production frequently represents a desirable recombinant protein production route.

57   Previous studies have attempted to increase the proportion of soluble recombinant protein

58   generated in bacterial hosts using a variety of methods. Some recombinant proteins can be induced

59   to fold using modulation of chaperones, a class of proteins that assist protein folding (Reviewed by

60   [10]). This can be done by individual chaperone overproduction or induction of the heat shock

61   response, the mechanism by which *E. coli* naturally responds to misfolded proteins; however, this

62   route is far from generic and needs to be individually tested and optimised for each recombinant

63   protein and host.

64   Another, more generically applicable, approach to increasing recombinant protein folding is

65   minimisation of stress, typified by reducing the rate of recombinant protein production thus allowing

66   recombinant protein translation and folding to proceed more slowly. Examples are the use of *E. coli*

67   strains that transcribe recombinant protein-encoding genes more slowly [11] and reduction of

68   inducer concentration and temperature [20]. In the latter system, two important improvements

69   were made to fermentation conditions using the IPTG-induced pET system [24] in order to produce

70   an aggregation-prone recombinant protein (CheY-GFP) in a soluble form. First, an IPTG

71 concentration of 8 μM was used instead of 0.5 mM, which decreased the rates of CheY-GFP

72 transcription and translation. Second, a growth temperature of 25 °C was used throughout the

73 process instead of growth at 37 °C before induction of recombinant protein production and 25 °C

74 after induction, in order to slow growth and translation rates and prevent cold shock. Although

75 these alterations decreased the rate of recombinant protein synthesis, the reduction in physiological

76 stress imposed on the bacteria meant that plasmid loss was greatly reduced, so far fewer bacteria

77 were non-productive, and the slower rate of synthesis meant that CheY-GFP could be folded into a

78 soluble form, decreasing IB formation. Overall, this 'improved' protocol resulted in far higher overall

79 recombinant protein yields per unit biomass and higher recombinant protein solubility.

80 In order that such methods are useful industrially, they need to be applied to high cell density

81 fermentation regimes such as fed-batch growth. In this study, we developed high cell density fed-

82 batch fermentations using stress-minimisation methods [20] in order to achieve four aims: high

83 biomass generation; high percentage of productive bacteria; high yield of recombinant protein; and

84 enhanced solubility of the generated recombinant protein. We used flow cytometry (FCM) as a

85 single-cell analysis tool to optimise the fermentations in terms of bacterial physiology and

86 productivity. Protocols were developed that tested industrially-derived semi-defined medium,

87 different carbon sources, and different points of induction of recombinant protein production.

88

89 **MATERIALS AND METHODS**

90 ***Bacterial strains, plasmids and microbiological methods***

91 *Escherichia coli* strain BL21* (DE3) (F⁻ *ompT hsdS*$_B$ (r$_B$⁻m$_B$⁻) *gal dcm rne131* λ(DE3)) was used

92 throughout (Invitrogen, Paisley, UK). The recombinant CheY-GFP fusion protein was encoded by the

93 pET20bhc-CheY-GFP plasmid [8,20], comprising the *E. coli cheY* gene fused to *gfp* cloned into

94 pET20bhc [7,27]. The *gfp* gene contains the S65T (maximum λ_{ex} red-shifted to 488 nm) and F64L

95 (folding improvement) mutations and so is equivalent to the GFPmut1 protein [4]. Bacteria were

96 transformed with the plasmid using the heat-shock method and transformants selected on nutrient

97 agar (Oxoid) plates supplemented with 100 µg carbenicillin mL^{-1} (Melford, Ipswich, UK; a more stable

98 variant of ampicillin). Optical density at 650 nm was used as a routine measurement technique for

99 biomass, due to its widespread use industrially and speed of data acquisition. For colony forming

100 unit (CFU) analysis, bacterial cultures were serially decimally diluted in PBS (8 $g \cdot L^{-1}$ NaCl, 0.2 $g \cdot L^{-1}$ KCl,

101 1.15 $g \cdot L^{-1}$ Na_2HPO_4, 0.2 $g \cdot L^{-1}$ KH_2PO_4, pH 7.3; Oxoid), plated onto nutrient agar plates (Oxoid) and

102 incubated at 25 °C for 48 hours. Colonies were replica plated onto plates containing 80 µg

103 carbenicillin mL^{-1} to determine plasmid retention. The dry cell weight (DCW) of four aliquots of 2 mL

104 of culture (harvested by centrifugation) was determined after drying at 100 °C for ≥24 hours.

105 ***Fermentation methods***

106 An Electrolab (Tewkesbury, UK) Fermac 310/60 5 L bioreactor was used with 4 baffles and an

107 agitator with 2 six-bladed Rushton turbines. Aeration was achieved by sparging air from below the

108 lower impeller at a rate of 3 $L \cdot min^{-1}$ through a reusable, autoclavable 0.22 µm filter (Sartorius).

109 Dissolved oxygen tension (DOT) was measured *in situ* using a D150 Oxyprobe (Broadley James) and

110 was maintained above a set point of 30% by increasing agitation to a maximum of 1000 RPM from a

111 minimum of 200-500 RPM. pH was measured by an F-695 Fermprobe (Broadley James) and was

112 controlled at a set point of 6.3 ± 0.1 with the automated addition of sterile 10% (v/v) NH_3 or 5% (v/v)

113 HCl. Off gas was passed through a condenser, autoclavable 0.22 µm filter (Sartorius), 2 catch pots

114 and analysed using a PrimaDB gas mass spectrometer (Thermo); data were logged and analysed

115 using GasWorks v1.0 (Thermo).

116 Inocula were grown from a sweep of cells from an agar plate in 35 mL of LB (10 $g \cdot L^{-1}$ Tryptone (BD

117 Bacto), 5 $g \cdot L^{-1}$ Yeast extract (BD Bacto), 5 $g \cdot L^{-1}$ NaCl (Sigma)) supplemented with 100 $µg \cdot mL^{-1}$

118   carbenicillin in a 250 mL conical flask, at 25°C and agitated at 150 RPM for 18-21 hours. Prior to

119   addition to the vessel, 5 mL of inoculum was removed and used for screening and analysis.

120   Five fermentation protocols were used as outlined in Table 1; the initial batch medium volume was

121   1.5 L in each case. Protocol A used a complex LB-based medium [20]: 10 g·L^{-1} Tryptone (BD Bacto), 5

122   g·L^{-1} Yeast extract (BD Bacto), 5 g·L^{-1} NaCl (Sigma), 1 mL·L^{-1} *E. coli* sulphur free salts and 1 mL·L^{-1}

123   Silicone antifoam (Corning); supplemented post-autoclaving with 5 g·L^{-1} glucose and 100 mg·L^{-1}

124   carbenicillin. *E. coli* sulphur-free salts comprised 8.2 g $MgCl_2.7H_2O$, 1 g $MnCl_2.4H_2O$, 0.4 g $FeCl_3.6H_2O$,

125   0.1 g $CaCl_2$ and 2 mL concentrated HCl in 100 mL of distilled water. CheY-GFP production was

126   induced by the addition of 8 µM IPTG at an OD_{650} of around 0.5. Five hours post-induction, 1 mM

127   Serine, 1 mM Threonine and 1 mM Asparagine were added to the bioreactor. The feed for protocol

128   A contained 100 g·L^{-1} tryptone, 50 g·L^{-1} yeast extract, 200 g·L^{-1} glucose, 10 mM serine, 10 mM

129   threonine, 10 mM asparagine, 100 mg.L^{-1} carbenicillin, 8µM IPTG, 1 mL·L^{-1} *E. coli* sulphur-free salts

130   and 0.1% (v/v) silicone antifoam in a final volume of 1 litre. Feeding began on depletion of initial

131   carbon source as indicated by an increase in the DOT (approximately 11 hours post-induction). The

132   feed rate initially was 13.69 mL·h^{-1} and was increased when on-line monitoring systems (DOT and

133   GC-MS) indicated that the feed rate had become growth-limiting (21.13 mL·h^{-1} at 30 hours, 27.17

134   mL·h^{-1} at 44.5 hours and 38.0 mL·h^{-1} 49.5 hours, feed was exhausted at approximately 57 hours [2]).

135   Protocols B-E used a semi-defined medium [30]: 14 g·L^{-1} $(NH_4)_2SO_4$, 20 g·L^{-1} yeast extract, 2 g·L^{-1}

136   KH_2PO_4, 16.5 g·L^{-1} K_2HPO_4, 7.5 g·L^{-1} citric acid, 1.5 mL·L^{-1} concentrated H_3PO_4 and 0.66 mL·L^{-1}

137   polypropylene glycol as antifoam; supplemented post-autoclaving with 34 mL·L^{-1} trace metal

138   solution, 10 mM $MgSO_4·7H_2O$, 2 mM $CaCl_2·2H_2O$ and 100 mg·L^{-1} carbenicillin. Protocols B-D used 35

139   g·L^{-1} Glycerol as a carbon source, protocol E used 5 g·L^{-1} glucose. Trace metal solution contained 3.36

140   g·L^{-1} $FeSO_4·7H_2O$, 0.84 g·L^{-1} $ZnSO_4·7H_2O$, 0.15 g·L^{-1} $MnSO_4·H_2O$, 0.25 g·L^{-1} $Na_2MoO_4·2H_2O$, 0.12 g·L^{-1}

141    $CuSO_4·5H_2O$, 0.36 g·L^{-1} H_3BO_3 and 48 mL·L^{-1} concentrated H_3PO_4. The feed composition and rate for

142 protocols B-E are described in Table 1. For protocol D, feeding began prior to depletion of carbon

143 source (17.5 hours post-induction), was paused to allow consumption of glycerol (18.5 hours) and

144 resumed once it was apparent that the glycerol had been consumed (22 hours). For protocol E, an

145 exponential feed profile was calculated using the following equation [23]:

$$F = \left(\frac{1}{S}\right) \times \left(\frac{\mu}{Y_{XS}} + m\right) \times X_0 \times e^{\mu t}$$

146 Where: F equals the feed rate into the bioreactor ($L \cdot h^{-1}$); X_0, total biomass in bioreactor at start of

147 feed (g dry cell weight); μ, specific growth rate set at 0.2 h^{-1}; t, time (h); S, feed glucose

148 concentration (400 $g \cdot L^{-1}$); Y_{xs}, cell yield on glucose (0.622 g biomass·g glucose^{-1} [28]); and m,

149 maintenance coefficient for glucose (0.00468 g glucose·g biomass$^{-1} \cdot h^{-1}$ [28]). A μ of 0.2 was chosen as

150 it is in the range of initial μ values observed (0.05-0.25) for glycerol fermentations in this study.

151 When the feed rate F reached 67.5 $mL \cdot h^{-1}$ (at approximately 25.6 hours post-induction), it was not

152 increased any further.

153 ***Flow cytometry***

154 Bacteria were analysed using a BD Accuri C6 flow cytometer (BD, Oxford, UK). Samples were excited

155 using a 488 nm solid state laser and fluorescence was detected using 533/30 BP (FL1 channel) and

156 670 LP (FL3 channel) filters corresponding to GFP and propidium iodide (PI) fluorescence

157 respectively. Bacteria were stained with PI to determine viability; PI can only enter dead bacteria. A

158 200 μg PI mL^{-1} stock solution was made up in distilled water and added to samples at a final

159 concentration of 4 μg PI mL^{-1}. Particulate noise was eliminated using a FSC-H threshold. 20 000 data

160 points were collected at a maximum rate of 2 500 events sec^{-1}. Data were analysed using CFlow (BD).

161 Percentages of GFP$^+$ (productive) bacteria were determined using a gate on a FSC-A versus FL1-A

162 intensity plot. Percentages of PI$^+$ (dead) bacteria were determined using a gate on a FL3-A versus

163 FL1-A intensity plot.

SDS-PAGE

165 Proteins were separated according to molecular weight using Tris/Glycine SDS-PAGE with a 15%

166 (w/v) polyacrylamide gel [18]. Bacterial cell pellets were suspended in sample buffer containing β-

167 mercaptoethanol and heated at 100 °C for 10 minutes before being loaded onto the gel. Equal

168 quantities of biomass were loaded into each lane. SDS-PAGE gels were stained with Coomassie Blue

169 and dried, then scanned (Canon Canoscan 9000F) and the density of each protein band quantified

170 using ImageJ [19] to permit calculation of the percentage of total protein that was CheY-GFP.

171 Independently, soluble and insoluble bacterial protein fractions were separated using BugBuster®

172 (Novagen). Bacterial cell pellets were suspended in a volume of BugBuster® equal to that of sample

173 buffer, incubated at room temperature for 10 minutes then fractionated by centrifugation at 16 873

174 g for 20 min. The pelleted insoluble fraction was subsequently washed in PBS to remove any residual

175 soluble protein. Both fractions were then resuspended in a volume of sample buffer equal to the

176 volume of BugBuster® used and incubated at 100 °C for 10 minutes before separation by SDS-PAGE

177 as above. This protocol results in soluble fractions that are twice the volume and hence half the

178 protein concentration of the insoluble; to ensure gels were loaded with samples from an equivalent

179 biomass twice the volume of soluble fractions were loaded on the gel. ImageJ was used to

180 determine the percentage of CheY-GFP in the soluble and insoluble fractions.

181 **RESULTS AND DISCUSSION**

182 This study investigated the production of the CheY-GFP fusion protein, a model 'difficult' protein

183 that is prone to misfolding and inclusion body formation when overexpressed in *E. coli* [8]. CheY is an

184 *E. coli* chemotaxis protein, and is fused here to the commonly-used *Aequorea victoria* green

185 fluorescent protein (GFP). Previous studies have demonstrated that GFP fluorescence correlates to

186 correct folding of the CheY-GFP fusion, so can be used as a measure of protein solubility and yield

187 [20,27]. The genes encoding CheY-GFP are carried by a pET vector [24]; expression is dependent

188   upon the IPTG-inducible T7 RNA polymerase gene encoded at the DE3 locus of the *E. coli* BL21* host.

189   Initial fermentations followed the 'improved' protocol described by Sevastsyanovich *et al*. [20],

190   referred to here as protocol A (Table 1). Growth data, measured using optical density at 650 nm,

191   reveal that growth proceeded for 48 hours post-induction (Figure 1a) up to an OD_{650} of around 71.

192   Feeding started 11 hours after induction, triggered by an increase in DOT (Supplemental figure 1)

193   indicating depletion of initial batch phase carbon source (5 $g \cdot L^{-1}$ glucose). Peak biomass

194   concentration as determined by dry cell weight analysis was 30.1 $g \cdot L^{-1}$.

195   Plasmid retention as determined by both replica plating and the GFP^+ phenotype of bacterial

196   colonies on agar plates remained above 94 and 97% respectively (Fig. 1b). We also used flow

197   cytometry (FCM [13]) to determine the green fluorescence of each bacterium; these data are

198   presented as the percentage of bacteria containing GFP (determined by applying a green

199   fluorescence/forward scatter gate, within which bacteria are considered to be GFP^+) and the mean

200   green fluorescence of GFP^+ cells, denoted FL1-A (Fig. 1c). The percentage of GFP^+ cells as determined

201   by FCM remained above 90% (Figure 1b), closely correlating with agar plate data. Less than 5% of

202   bacteria were dead throughout the fermentation as determined by FCM and Propidium Iodide (PI)

203   staining. Online gas MS data (Supplementary Figure 1) reveal that oxygen demand and CO_2 evolution

204   rates dropped sharply after around 55 hours post-induction, corresponding to the end of the feed.

205   The mean green fluorescence of CheY-GFP^+ bacteria (FL1-A, determined using FCM) increased over

206   the course of the fermentation following a small decrease immediately post-induction (Fig. 1c),

207   thought to be caused by a concurrent decrease in bacterial size (as indicated using FCM forward

208   scatter [FSC-A] measurements). The cell size dynamics observed here are concomitant with previous

209   studies of *E. coli* cell size over growth curves; an increase in cell size during lag phase, followed by a

210   decrease in cell size during exponential growth [1].

211 Maximum green fluorescence per bacterium did not increase significantly after 28 hours post-

212 induction, although biomass increased thus increasing the quantity of recombinant protein present

213 in the fermenter as a whole. SDS-PAGE analysis of whole bacteria and soluble and insoluble bacterial

214 fractions (Fig. 1d) revealed that CheY-GFP concentration per unit biomass (expressed as a

215 percentage of whole cell protein) did not dramatically change during the fermentation, increasing

216 slightly from 11.4 % at induction to 16.8 % at 24 hours post-induction, then fluctuated over the

217 remainder of the fermentation. The percentage of CheY-GFP in the soluble fraction (as determined

218 by Bugbuster® extraction) did increase from around 40% soluble at induction to over 60% soluble 26

219 hours post induction; this increase occurred at the same time as the increase in green fluorescence

220 as measured by FCM.

221 Taken together, the protocol A fermentation could therefore be split into three phases: 0-28 hours

222 post-induction, concurrent increase in biomass and quantity of recombinant protein per cell; 28-48

223 hours, biomass accumulation but no change in quantity of CheY-GFP per cell; and accumulation of

224 neither biomass nor recombinant protein after 48 hours post-induction. 48 hours could be

225 determined to be the optimal harvest time. The final yield of total CheY-GFP produced (at 70 hours

226 post-induction) was estimated to be 2.3 $g \cdot L^{-1}$ (assuming protein comprises 60 % of dry cell weight,

227 based on observations of 50-61 % [26] and an estimate of 70 % by Sevastsyanovich *et al*. [20]),

228 corresponding to a yield of 1.5 $g \cdot L^{-1}$ soluble CheY-GFP.

229 Three potential problems were identified with protocol A [20] that could limit its utility in

230 biopharmaceutical manufacturing. First, both the base medium and feed contain complex animal-

231 derived products (tryptone) so are not suitable for a cGMP process. Many industrial RPP processes

232 tend to use defined media or semi-defined media without animal products; choice of defined or

233 semi-defined media is usually down to company policy and product. The use of semi-defined media

234 offers a compromise before the potentially high cost of development of a fully defined medium

235 optimised for a particular bioprocess. Minimisation of the use of complex media components also

236 decreases the risks of batch variability. This variability was characterised for this system by growing

237 the *E. coli* BL21* pCheY-GFP strain for 14 hours at 30 °C in LB medium composed of complex medium

238 components sourced from different suppliers: a variability of 21 % in final OD_{650} was observed due to

239 differences in yeast extract and tryptone composition.

240 Second, large quantities of additional complex medium components are fed into the fermenter

241 during the fed-batch phase. This can result in osmotic problems and the presence of large quantities

242 of undefined proteinaceous medium components can complicate downstream processing of product

243 proteins [5]. Finally, use of glucose as a feed can present difficulties from acid formation due to

244 overflow metabolism, especially when growth rates fluctuate. Use of glycerol as a carbon source

245 does not usually present this problem. These three concepts in industrial fermentation design are

246 typified by the protocol used by Want *et al*. [30], which was used to test CheY-GFP expression here

247 (referred to as protocol B; Table 1).

Use of an industrially-derived fermentation protocol

249 As well as the use of semi-defined medium and glycerol as carbon source, protocol B used a growth

250 temperature of 37 °C and induction at a relatively high biomass using a high concentration (0.1 mM)

251 of IPTG. The glycerol feed was started at the same time as induction, when online measurements

252 suggested exhaustion of batch-phase glycerol (primarily by reduction in oxygen demand), and was

253 fed at a rate of 67.5 mL·h^{-1}. This protocol resulted in more heterogenous data than protocol A; data

254 from two fermentations are shown in Figure 2. The growth data in terms of OD_{650} and CFU

255 measurements are similar for both fermentations (Fig. 2a). Cell density increased to an OD_{650} of

256 around 55, whereupon RPP was induced by addition of 0.1 mM IPTG. Growth continued until an

257 OD_{650} of around 80; oxygen consumption data (DOT and gas MS) reveal that induction caused

258 growth arrest followed by a recovery, but rapid growth only proceeded for around 5 hours post-

11

259 induction (Supplementary figure 2). This is indicative of metabolic stress generated upon CheY-GFP

260 synthesis at 37 °C [20].

261 Although the growth of duplicate fermentations was similar, CheY-GFP production in each

262 fermentation was quite different. Both replicates showed high levels of plasmid loss before

263 induction (as determined both by plating and FCM; Figure 2b), suggesting that even uninduced cells

264 were under stress. CheY-GFP is synthesised from this plasmid system even in the absence of inducer

265 (Fig. 1), thus imposing stress before induction [20]. FCM data revealed that 15 hours post-induction,

266 >30% of bacteria were still CheY-GFP$^+$ in fermenter 1; despite this, on agar plates, all colonies from

267 fermenter 1 were GFP$^-$ at this timepoint. This is likely caused by physiological stress in CheY-GFP$^+$

268 bacteria generating a viable but non-culturable (VBNC) phenotype, commonly encountered in

269 bacterial recombinant protein production cultures. Plasmid loss was greater in fermentation 2 than

270 1; only ≈10% of bacteria were CheY-GFP$^+$. The proportion of dead (PI$^+$) bacteria in fermenter 1 was

271 also higher than in fermenter 2, and the CFU·mL^{-1} was lower. Mean green fluorescence of the CheY-

272 GFP$^+$ bacteria were equivalent (Fig. 2c), but the much lower proportion of productive bacteria in

273 fermenter 2 meant that far less CheY-GFP was produced per unit biomass (Fig. 2d). It is interesting to

274 note that the solubility of CheY-GFP produced by fermenter 2 was very high; this is probably a

275 consequence of the very low quantity of CheY-GFP being produced.

276 For fermenter 1, the final yield of CheY-GFP was estimated at 2.2 g·L^{-1}, corresponding to a yield of

277 0.7 g·L^{-1} soluble CheY-GFP. It can be concluded that fermenter 1 was more productive in terms of

278 CheY-GFP productivity per unit biomass, but that bacteria were under greater physiological stress.

279 Fermenter 2 had a lower proportion of productive bacteria, resulting in a lower CheY-GFP yield (≈

280 0.6 g·L^{-1} CheY-GFP, ≈ 0.5 g·L^{-1} soluble CheY-GFP) but less physiological stress, resulting in higher CFU

281 measurements and a lower proportion of dead bacteria.

282　The low overall levels of plasmid retention, green fluorescence and CheY-GFP accumulation and

283　solubility as compared to protocol A suggests that protocol B put bacteria under physiological stress

284　that was detrimental to recombinant protein production. Therefore, the improvements used to

285　initially design the improved protocol A were applied to the industrially-derived protocol B; namely

286　reduction of temperature to 25 °C throughout and induction with a far lower concentration of IPTG

287　(8 μM) to generate protocol C (Table 1).

Application of improved conditions to an industrial protocol

289　Due to the low growth temperature compared to protocol B, protocol C cultures grew far slower

290　(Fig. 3a), taking around 21 hours to reach the induction point (OD_{650} ≈ 40, when glycerol was

291　exhausted as indicated by online measurements). Following induction, the culture grew well for 11

292　hours as indicated by a steady increase in oxygen demand (DOT and gas MS data; Supplementary

293　figure 3). The feed likely became growth limiting at 11 hours post-induction, even though it lasted

294　until around 18 hours post induction, at which point oxygen demand (as determined by gas MS) fell

295　dramatically. The final OD_{650} was recorded as 297. Unlike protocol B cultures, the proportion of GFP+

296　bacteria as measured by FCM remained very high throughout (>93%), indicative of good plasmid

297　retention and a consequence of lowered physiological stress (Fig. 3b). Mean green fluorescence per

298　GFP+ bacterium as determined by FCM (FL1-A) decreased between inoculation and induction

299　(concurrent with a decrease in cell size, signified by mean FSC-A measurements) and, as in protocol

300　A, increased following induction, although this increase was to a far greater extent than protocol A

301　cultures, reaching a peak mean green fluorescence of 240 000 compared to 176 000 for protocol A

302　(Fig. 3c). Unlike protocol A cultures, mean forward scatter (FSC-A), signifying bacterial size, increased

303　following induction, and was greater at the end of the fermentation in protocols B and C than in

304　protocol A. This is probably due to the higher osmolarity in the medium and feed used in protocol A

305　[16].

306   CheY-GFP solubility as determined by Bugbuster® peaked at nearly 60% at the point of induction,

307   then decreased to a low point at 5 hours post-induction however recovered to almost the peak value

308   at termination reaching a final solubility of around 56% (Fig. 3d). Total CheY-GFP accumulation per

309   unit biomass followed a similar pattern however peak accumulation occurred at 2 hours post-

310   induction. These data suggest that CheY-GFP concentration per unit biomass and solubility were

311   lowest during periods of rapid growth. The final yield of CheY-GFP was estimated at 10.7 $g \cdot L^{-1}$,

312   corresponding to a yield of 6 $g \cdot L^{-1}$ soluble CheY-GFP.

313   Summary data comparing protocols B and C (Table 2) clearly demonstrate the benefits of operating

314   using stress minimisation conditions in terms of the resultant biomass yield (Final OD_{650} increased

315   nearly 4-fold; DCW increased almost 3-fold) and recombinant protein yield and solubility. Compared

316   to protocol A, Protocol C showed an almost 5-fold increase in cell density and although FCM analysis

317   showed similar percentages of GFP^+ cells, suggesting similar levels of plasmid retention, GFP^+

318   bacteria in protocol C at harvest were 46 % more fluorescent than in protocol A and showed higher

319   levels of homogeneity as evidenced by the lower CV of the green fluorescence values. CheY-GFP

320   solubility as assessed by BugBuster® fractionation was 11% higher in protocol A than in protocol C;

321   this is in agreement with Moore *et al*. [12] who showed that increasing the concentration of complex

322   media components (tryptone and yeast extract) increased solubility of recombinant T4 dCMP

323   deaminase*.* However, CheY-GFP concentration as a percentage of total cellular protein in protocol C

324   was almost double that of protocol A; this may partially explain the slightly lower solubility, the

325   higher quantity of CheY-GFP having overwhelmed the bacterial protein folding pathways. These

326   data, combined with the increased biomass, resulted in an over 4-fold increase in CheY-GFP

327   volumetric yield and a 4-fold increase in the volumetric yield of soluble CheY-GFP.

328   ***Alteration of induction point***

14

329 As in protocol B [30], many RPP protocols induce recombinant protein production at a relatively high

330 biomass in order to separate biomass generation and recombinant protein production stages. This is

331 often done when the recombinant protein in question is a 'toxic' protein, prone to cause host

332 bacteria stress; in addition this can lead to reduction in metabolic burden that can be caused by

333 simultaneous requirements for cellular resources for both biomass and recombinant protein

334 generation. For many recombinant protein production processes, the time from induction to time

335 for harvest is limited (the 'production window') and is governed by the amount of time that the host

336 bacteria can generate recombinant protein without losing viability. Protocol A utilised early

337 induction, using the logic that reduction in stress and lower growth rates would allow bacteria to

338 generate recombinant protein more slowly and thus apportion cellular resources more evenly

339 between biomass and recombinant protein generation. Therefore, protocol C was modified to allow

340 induction at an earlier point ($OD_{650} \approx 0.5$), as in protocol A, to generate protocol D.

341 Cell growth was broadly comparable to protocol C, taking around 40 hours to reach an OD_{650} of 288

342 (Fig. 4a). Online data revealed that metabolic activity declined at 28 hours post-induction

343 (Supplementary Fig. 4), reflecting a decrease in growth rate; as with protocol C, this is probably due

344 to the feed rate limiting growth. The proportion of GFP[+] cells as determined by FCM (Fig. 4b) during

345 early stages of the fermentation was lower than expected at approximately 80%; however, this was

346 mainly due to non-fluorescent antifoam particulate matter with a similar scatter distribution to

347 bacteria; this particulate noise was visible due to a low cell density at the start of these

348 fermentations. At later points in the fermentation the percentage of GFP[+] cells remained above 95%

349 but at termination the proportion had dropped to 74%. PI staining also showed an increase in the

350 percentage of dead cells at termination up to 8.8%. These data suggest that by termination the

351 culture had become physiologically stressed. As before, the mean green fluorescence of GFP[+]

352 bacteria and the mean forward scatter initially decreased (Fig 4c); after 18 hours post-induction both

353 parameters steadily increased until termination, FL1-A reaching a final value of 370 000. The

354 increase in FL1-A during the latter stages of the fermentation was greater than that of FSC-A

355 (signifying cell size), suggesting accumulation of CheY-GFP per bacterium. FSC-A changes were

356 similar to protocol C, except that induction occurred at the maximum mean FSC-A value rather than

357 the minimum; this suggests that bacterial size is primarily regulated in response to growth phase and

358 is not a result of recombinant protein production. SDS-PAGE analysis (Figure 4d) showed a steady

359 increase in the percentage of total cellular protein that was CheY-GFP throughout; from 16% at the

360 point of induction to 26% at termination. However the percentage solubility showed an overall

361 decrease during the fermentation from 47% at induction to 37% at termination, suggesting that

362 overall product quality had decreased; again, possibly due to higher rates of CheY-GFP synthesis, as

363 seen in protocol C. Final yield of CheY-GFP was estimated at 12 $g \cdot L^{-1}$, corresponding to a yield of 4.5

364 $g \cdot L^{-1}$ soluble CheY-GFP.

365 Comparing protocols C and D allow the effect of early versus late induction to be examined. In terms

366 of biomass generation protocols C and D showed similar final measurements (Table 2). In addition,

367 culture viability, as indicated by the percentage of PI^+ cells, was similar. It can therefore reasonably

368 be concluded that culture growth and biomass generation did not appear to be affected by earlier

369 induction. In terms of CheY-GFP productivity the effect of early induction was an increase in

370 heterogeneity within the culture. The percentage of total protein that was CheY-GFP was similar at

371 harvest, but the product quality, as indicated by CheY-GFP solubility, was almost 20% lower in

372 protocol D. The proportion of GFP^+ cells was over 20% lower for protocol D, but the mean green

373 fluorescence of the GFP^+ cells was over 50% higher. Nonetheless, similar amounts of CheY-GFP per

374 unit biomass were observed by SDS-PAGE in protocols C and D, although its solubility was lower in

375 protocol D.

376 Based on these data, earlier induction increased culture heterogeneity in protocol D, evidenced by a

377 larger number of GFP^- cells, and a lower percentage of soluble CheY-GFP as determined by SDS-

16

378 PAGE. It is possible that induction of RPP tends to select for culture heterogeneity, even in stress-

379 minimising conditions; the longer time between induction and harvest allowed a larger

380 subpopulation to develop in protocol D than C. This represents an additional factor when choosing a

381 harvest window.

382 ***Use of glucose as a carbon source***

383 Although glycerol has advantages over glucose as a carbon source, it is more expensive and the

384 preferred carbon and energy source of *E. coli* is glucose. Therefore, protocol D was modified to use

385 glucose as a carbon source both in the batch medium and the feed, generating protocol E. Growth

386 data reveal steady growth to a final OD_{650} of 167 (Fig. 5a); gas MS revealed that growth significantly

387 slowed at 32 hours post-induction (Supplemental Figure 5), corresponding to the end of the glucose

388 feed. Plasmid retention remained above 92% throughout (Fig. 5b). The percentage of GFP$^+$ cells

389 determined by FCM remained above 98%, except for the initial sample that was 85%, again caused

390 by antifoam particulates. The percentage of dead cells as determined by FCM remained at less than

391 7% throughout, although there was an increase between 26 hours (1.8%) and termination (5.9%),

392 possibly suggesting a increase in cell stress due to the onset of stationary phase.

393 Mean green fluorescence and forward scatter of GFP$^+$ cells (Fig. 5c) showed a similar pattern to

394 other early-induced fermentations. Between 6 and 20 hours post-induction FSC-A decreased to a

395 greater extent than FL1-A, suggesting that while the cells became smaller, CheY-GFP content per cell

396 increased. Between 20 and 26 hours post-induction FL1-A increased by approximately 30% and FSC-

397 A increased by approximately 20%, suggesting accumulation of CheY-GFP per cell despite increasing

398 cell size. At termination mean green fluorescence per bacterium had reached 296 000. SDS-PAGE

399 data (Figure 5d) showed an increase in the percentage of total cellular protein that was CheY-GFP

400 from 2 hours post-induction until termination, reaching a final peak value of 30%. CheY-GFP

401 solubility fluctuated during the fermentation, the peak solubility of 50% being observed at 5 hours

402   post-induction and final solubility being 46%. Final yield of CheY-GFP was estimated at 7.7 $g \cdot L^{-1}$,

403   corresponding to a yield of 3.8 $g \cdot L^{-1}$ soluble CheY-GFP.

404   Although utilising the same carbon source, protocol E shows several benefits over protocol A: a 2.5-

405   fold increase in OD_{650}; a 65% increase in the mean green fluorescence of GFP^+ bacteria; and the

406   largest amount of CheY-GFP as expressed as a percentage of cellular protein achieved in this study.

407   Although the soluble percentage of CheY-GFP was higher in protocol A than E, the far higher

408   quantity of CheY-GFP per cell and higher biomass concentration meant that the volumetric yield of

409   soluble CheY-GFP in protocol E was over double that of protocol A (3.8 $g \cdot L^{-1}$ versus 1.5 $g \cdot L^{-1}$). As

410   before, increases in CheY-GFP per cell probably correlate with decreases in CheY-GFP solubility due

411   to overloading of cellular protein folding pathways.

412   Comparison of protocols E and D allows elucidation of differences caused by changing carbon

413   source. There are numerous studies that claim recombinant protein production is enhanced by

414   growth on either glucose [3,25] or glycerol [9,15,31]; this seems to be dependent upon recombinant

415   protein, strain, medium composition and growth conditions, and so is likely not a generic effect. In

416   this study, protocol E had a lower final OD_{650} than D; however, this was expected as total glucose

417   added was less than glycerol, and $Y_{X/S}$ values were broadly comparable (Table 2). FCM data

418   demonstrate that there were fewer GFP^- bacteria in protocol E than D, suggesting lower

419   physiological stress, although protocol D had a longer runtime than E, which could select for a non-

420   productive GFP^- subpopulation. CheY-GFP yield per cell and solubility were higher in protocol E,

421   although the higher biomass generated in protocol D resulted in a higher overall CheY-GFP yield.

422   Again, there is a balance between biomass production, CheY-GFP production and solubility;

423   protocols C-E have demonstrated that although each of these parameters may be optimised

424   individually, it is at the expense of other parameters.

425   **CONCLUSIONS**

426   The stress minimisation method [20] has been shown to be highly applicable to an industrially-

427   derived high cell density fed-batch recombinant protein production protocol, both with early and

428   late induction of RPP and with glucose and glycerol as carbon sources. Stress minimisation increased

429   biomass yield and CheY-GFP yield and solubility while decreasing culture heterogeneity. Similarly,

430   transfer to a semi-defined medium improved biomass yield and overall CheY-GFP productivity per

431   unit volume, while representing a more industrially-favoured approach to RPP due to elimination of

432   animal-derived products and minimisation of complex media components. Changing the point of

433   induction was shown to have little overall effect on the improved protocols. Flow cytometry was

434   shown to be a very useful analytical tool in fermentation monitoring and optimisation, in particular

435   allowing culture heterogeneity, stress and the relationship between bacterial size and GFP content

436   to be monitored. In summary, the stress minimisation methods described here could effectively be

437   applied to a wide range of high cell density culture recombinant protein production fermentations.

442   **CONFLICT OF INTEREST STATEMENT**

443   The BD Accuri C6 flow cytometer was awarded to TWO by the BD Accuri Creativity Award. TWO was

444   paid speakers expenses by BD for speaking at a BD Accuri users' event. The funders played no role in

445   the design or implementation of this study.

446   **REFERENCES**

447   1.   Åkerlund T, Nordström K, Bernander R (1995) Analysis of cell size and DNA content in exponentially growing and stationary-phase batch cultures of *Escherichia coli*. J Bacteriol 177: 6791–6797

450   2.   Alfasi S, Sevastsyanovich Y, Zaffaroni L, Griffiths L, Hall R, Cole J (2011) Use of GFP fusions for the isolation of *Escherichia coli* strains for improved production of different target recombinant proteins. J Biotechnol 156: 11-21.

453   3.   Carvalho RJ, Cabrera-Crespo J, Tanizaki MM, Goncalves VM (2012) Development of production and purification processes of recombinant fragment of pneumococcal surface protein A in *Escherichia coli* using different carbon sources and chromatography sequences. Appl Microbiol Biotechnol 94: 683-94.

457   4.   Cormack BP, Valdivia RH, Falkow S (1996) FACS-optimized mutants of the green fluorescent protein (GFP). Gene. 173: 33-8.

459   5.   Cote RJ (2010) Media composition, microbial, laboratory scale. In: Encyclopedia of Industrial Biotechnology, Bioprocess, Bioseparation, and Cell Technology. Flickinger MC, ed. John Wiley & Sons. pp 3277-3297

462   6.   García-Fruitós E, González-Montalbán N, Morell M, Vera A, Ferraz RM, Arís A, Ventura S, Villaverde A (2005) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. Microb Cell Fact 4: 27

465   7.   Jones JJ, Bridges AM, Fosberry AP, Gardner S, Lowers RR, Newby RR, James PJ, Hall RM, Jenkins O (2004) Potential of real-time measurement of GFP-fusion proteins. J Biotechnol 109: 201-11

467   8.   Jones JJ (2007) Green fluorescent protein as an analytical tool to dissect the physiology of recombinant protein production in fermenters. PhD thesis, University of Birmingham, Birmingham, UK.

470   9.   Luo Q, Shen YL, Wei DZ, Cao W (2006) Optimization of culture on the overproduction of TRAIL in high-cell-density culture by recombinant *Escherichia coli*. Appl Microbiol Biotechnol 71: 184-91.

472   10. Martínez-Alonso M, García-Fruitós E, Ferrer-Miralles N, Rinas U, Villaverde A (2010) Side effects

473   of chaperone gene co-expression in recombinant protein production. Microb Cell Fact 9: 64.

474   11. Miroux, B and Walker, JE (1996) Over-production of proteins in Escherichia coli: mutant hosts

475   that allow synthesis of some membrane proteins and globular proteins at high levels. J Mol Biol

476   260: 289-98.

477   12. Moore JT, Uppal A, Maley F, Maley GF (1993) Overcoming inclusion body formation in a high-

478   level expression system. Protein Expr Purif 4: 160-3.

479   13. Müller S, Nebe-von-Caron G (2010) Functional single-cell analyses: flow cytometry and cell

480   sorting of microbial populations and communities. FEMS Microbiol Rev 34: 554-87.

481   14. Overton TW (2014) Recombinant protein production in bacterial hosts. Drug Discov Today *In*

482   *press* doi:10.1016/j.drudis.2013.11.008

483   15. Pflug, S, Richter, S M & Urlacher, V B (2007) Development of a fed-batch process for the

484   production of the cytochrome P450 monooxygenase CYP102A1 from *Bacillus megaterium* in *E.*

485   *coli*. J Biotechnol 129: 481-8.

486   16. Reuter M, Hayward NJ, Black SS, Miller S, Dryden DT, Booth IR (2013) Mechanosensitive

487   channels and bacterial cell wall integrity: does life end with a bang or a whimper? J R Soc

488   Interface. 11: 20130850. doi: 10.1098/rsif.2013.0850.

489   17. Saïda F, Uzan M, Odaert B, Bontems F (2006) Expression of highly toxic genes in *E. coli*: special

490   strategies and genetic tools. Curr Protein Pept Sci 7: 47-56.

491   18. Sambrook J, Fritsch EF, Maniatis T (1987) Molecular Cloning: A Laboratory Manual, 2nd edn.

492   Cold Spring Harbor Press, Cold Spring Harbor, NY.

493   19. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis.

494   Nat Methods 9: 671-675.

495   20. Sevastsyanovich Y, Alfasi S, Overton T, Hall R, Jones J, Hewitt C, Cole J (2009) Exploitation of GFP

496   fusion proteins and stress avoidance as a generic strategy for the production of high-quality

497   recombinant proteins. FEMS Microbiol Lett 299: 86–94.

498   21. Singh, SM and Panda, AK (2005) Solubilization and refolding of bacterial inclusion body proteins.

499   J Biosci Bioeng 99: 303-10.

500   22. Sørensen HP, Mortensen KK. (2005) Soluble expression of recombinant proteins in the

501   cytoplasm of *Escherichia coli*. Microb Cell Fact. 4: 1.

502   23. Strandberg, L, Andersson, L, Enfors, S (1994) The use of fed batch cultivation for achieving high

503   cell densities in the production of a recombinant protein in Escherichia coli. FEMS Microbiol Rev

504   14: 53–6.

505   24. Studier FW(1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression

506   system. J Mol Biol 219: 37-44.

507   25. Tseng, C L & Leng, C H (2012) Influence of medium components on the expression of

508   recombinant lipoproteins in *Escherichia coli*. Appl Microbiol Biotechnol 93: 1539-52.

509   26. Valgepea, K, Adamberg, K, Seiman, A, Vilu, R (2013) *Escherichia coli* achieves faster growth by

510   increasing catalytic and translation rates of proteins. Mol Biosyst 9: 2344-58.

511   27. Waldo GS, Standish BM, Berendzen J, Terwilliger TC (1999) Rapid protein-folding assay using

512   green fluorescent protein. Nat Biotechnol 17: 691-5.

513   28. Wallace, R, Holms, W (1986) Maintenance coefficients and rates of turnover of cell material in

514   *Escherichia coli* ML308 at different growth temperatures. FEMS Microbiol Lett 37: 317-320

515   29. Walsh G (2010) Biopharmaceutical benchmarks. Nat. Biotechnol 28: 917-24.

516   30. Want A, Thomas OR, Kara B, Liddell J, Hewitt CJ. (2009) Studies related to antibody fragment

517   (Fab) production in *Escherichia coli* W3110 fed-batch fermentation processes using

518   multiparameter flow cytometry. Cytometry A 75: 148-54.

519   31. Zhang JD, Li AT, Xu JH (2010) Improved expression of recombinant cytochrome P450

520   monooxygenase in *Escherichia coli* for asymmetric oxidation of sulfides. Bioprocess Biosyst Eng

521   33: 1043-9.

522

523   **FIGURE LEGENDS**

524   **Figure 1.** Fermentation data using protocol A (Original 'improved' protocol, complex medium and

525   feed, early induction). A. OD_{650} (squares) and CFU (crosses) data. Arrow indicates time of feed

526   starting. B. Plasmid retention data; percentage of colonies that were GFP[+] (squares) and percentage

527   plasmid[+] colonies (crosses) as determined using agar plates. Flow cytometry data; percentage of

528   cells that were GFP[+] (circles) and percentage of cells that were PI[+] and so dead (triangles). C. Flow

529   cytometry data; mean cellular green fluorescence (FL1-A) of GFP[+] cells (squares) and mean forward

530   scatter (FSC-A) of GFP[+] cells (crosses). D. SDS-PAGE data; percentage solubility of CheY-GFP (squares)

531   and percentage of total protein that was CheY-GFP (crosses). Data from a representative

532   fermentation of a minimum of 2 replicates.

533   **Figure 2.** Fermentation data using protocol B (Industrially derived 'standard' conditions, semi-

534   defined medium, late induction, high [IPTG]). A - OD_{650} data for replicate 1 (squares) and 2 (crosses)

535   CFU data for replicate 1 (circles) and 2 (triangles). Arrow indicates time of feed starting. B - Plasmid

536   retention data; percentage of colonies that were GFP[+] (squares) and percentage plasmid[+] colonies

537   (crosses) as determined using agar plates. Flow cytometry data; percentage of cells that were GFP[+]

538   (circles) and percentage of cells that were PI[+] and so dead (triangles). Replicate 1 – left, replicate 2 –

539   right. C - Flow cytometry data; mean cellular green fluorescence (FL1-A) of GFP[+] cells (squares) and

540   mean forward scatter (FSC-A) of GFP[+] cells (crosses). Replicate 1 – left, replicate 2 – right. D - SDS-

541   PAGE data; percentage solubility of CheY-GFP (squares) and percentage of total protein that was

542   CheY-GFP (crosses). Replicate 1 – left, replicate 2 – right.

543 **Figure 3.** Fermentation data using protocol C (Improved industrially derived conditions, semi-defined

544 medium, late induction, low [IPTG]). A. OD_{650} (squares) and CFU (crosses) data. Arrow indicates time

545 of feed starting. B. Flow cytometry data; percentage of cells that were GFP[+] (squares) and

546 percentage of cells that were PI[+] and so dead (crosses). C. Flow cytometry data; mean cellular green

547 fluorescence (FL1-A) of GFP[+] cells (squares) and mean forward scatter (FSC-A) of GFP[+] cells (crosses).

548 D. SDS-PAGE data; percentage solubility of CheY-GFP (squares) and percentage of total protein that

549 was CheY-GFP (crosses). Data from a representative fermentation of a minimum of 2 replicates.

550 **Figure 4.** Fermentation data using protocol D (Improved industrially derived conditions, semi-

551 defined medium, early induction, low [IPTG]). A. OD_{650} (squares) data. Arrow indicates time of feed

552 starting. B. Flow cytometry data; percentage of cells that were GFP[+] (squares) and percentage of

553 cells that were PI[+] and so dead (crosses). C. Flow cytometry data; mean cellular green fluorescence

554 (FL1-A) of GFP[+] cells (squares) and mean forward scatter (FSC-A) of GFP[+] cells (crosses). D. SDS-PAGE

555 data; percentage solubility of CheY-GFP (squares) and percentage of total protein that was CheY-GFP

556 (crosses). Data from a representative fermentation of a minimum of 2 replicates.

557 **Figure 5.** Fermentation data using protocol E (Improved industrially derived conditions, semi-defined

558 medium, early induction, low [IPTG], glucose feed). A. OD_{650} (squares) and CFU (crosses) data. Arrow

559 indicates time of feed starting. B. Plasmid retention data; percentage of colonies that were GFP[+]

560 (squares) and percentage plasmid[+] colonies (crosses) as determined using agar plates. Flow

561 cytometry data; percentage of cells that were GFP[+] (circles) and percentage of cells that were PI[+] and

562 so dead (triangles). C. Flow cytometry data; mean cellular green fluorescence (FL1-A) of GFP[+] cells

563 (squares) and mean forward scatter (FSC-A) of GFP[+] cells (crosses). D. SDS-PAGE data; percentage

564 solubility of CheY-GFP (squares) and percentage of total protein that was CheY-GFP (crosses). Data

565 from a representative fermentation of a minimum of 2 replicates.

	A – Original improved protocol	B – Industrially derived, standard conditions	C – Industrially derived, improved conditions	D – Industrially derived, improved conditions, mid logarithmic induction, glycerol	E – Industrially derived, improved conditions, mid logarithmic induction, glucose
Growth medium	Complex [20]	Semi-defined [30]	Semi-defined [30]	Semi-defined [30]	Semi-defined [30]
Carbon source in batch phase	5 g·L^{-1} glucose	35 g·L^{-1} glycerol	35 g·L^{-1} glycerol	35 g·L^{-1} glycerol	5 g·L^{-1} glucose
Feed composition and volume	1 L; 100 g·L^{-1} Tryptone 50 g·L^{-1} Yeast extract 200 g·L^{-1} Glucose 10 mM each serine, threonine, asparagine 100 mg.L^{-1} carbenicillin, 8µM IPTG, 1 mL·L^{-1} *E. coli* sulphur-free salts 0.1% (v/v) silicone antifoam	0.5 L 714 g·L^{-1} glycerol 7.4 g·L^{-1} MgSO$_4$·7H$_2$O	0.5 L 714 g·L^{-1} glycerol 7.4 g·L^{-1} MgSO$_4$·7H$_2$O	0.5 L 714 g·L^{-1} glycerol 7.4 g·L^{-1} MgSO$_4$·7H$_2$O	0.5 L 400 g·L^{-1} glucose 7.4 g·L^{-1} MgSO$_4$·7H$_2$O
Feed rate	Stepped (*see Methods*)	67.5 mL·h^{-1}	67.5 mL·h^{-1}	67.5 mL·h^{-1}	Exponential at µ = 0.2, up to 67.5 mL·h^{-1}
Feed start	On glucose exhaustion	OD$_{650}$ ≈ 40-50	OD$_{650}$ ≈ 40-50	On glycerol exhaustion	On glucose exhaustion
Temperature	25°C	37°C	25°C	25°C	25°C
[IPTG]	8 µM	100 µM	8 µM	8 µM	8 µM
Induction point	Mid-logarithmic phase (OD$_{650}$ ≈ 0.5)	With feeding (OD$_{650}$ ≈ 40-50)	With feeding (OD$_{650}$ ≈ 40-50)	Mid-logarithmic phase (OD$_{650}$ ≈ 0.5)	Mid-logarithmic phase (OD$_{650}$ ≈ 0.5)

Table 1: Summary of fermentation protocols

Protocol	Peak / final[a] OD$_{650}$	DCW (g·L^{-1})	Y$_{xs}$ (g·g^{-1})	Flow cytometry measurements						SDS-PAGE		Total CheY::GFP (g·L^{-1})[c]	Total soluble CheY::GFP (g·L^{-1})
				% bacteria GFP$^+$	% bacteria PI$^+$	Mean FL1-A (GFP$^+$ bacteria)	Mean FL1-A (all bacteria)	% CV (Coefficient of Variation) FL1-A	Mean FSC-A (GFP$^+$ bacteria)	% CheY::GFP soluble	% of protein that is CheY::GFP		
A	64.5 (71.8)	30.1	0.363	94.3	2.1	166 700 (176 100)	157 200	73	8 800	62.5	13.0	2.3	1.5
B[b]	72.4 (81.8)	27.9	0.136	33.0	6.7	117 700 (117 800)	62 800	93	29 100	30.0	13.4	2.2	0.7
C	297	72.8	0.365	95.2	9.3	240 100	228 600	54	23 900	55.6	24.5	10.7	6
D	288	77.5	0.378	73.5	8.8	369 400	271 600	83	25 200	37.4	25.7	12.0	4.5
E	167	42.6	0.411	98.2	5.9	295 500	290 100	73	22 700	49.5	30.1	7.7	3.8

Table 2. Summary of fermentation data at harvest.

[a] Values in parentheses are peak measurements.

[b] Values refer to fermentation 1 (Fig. 2)

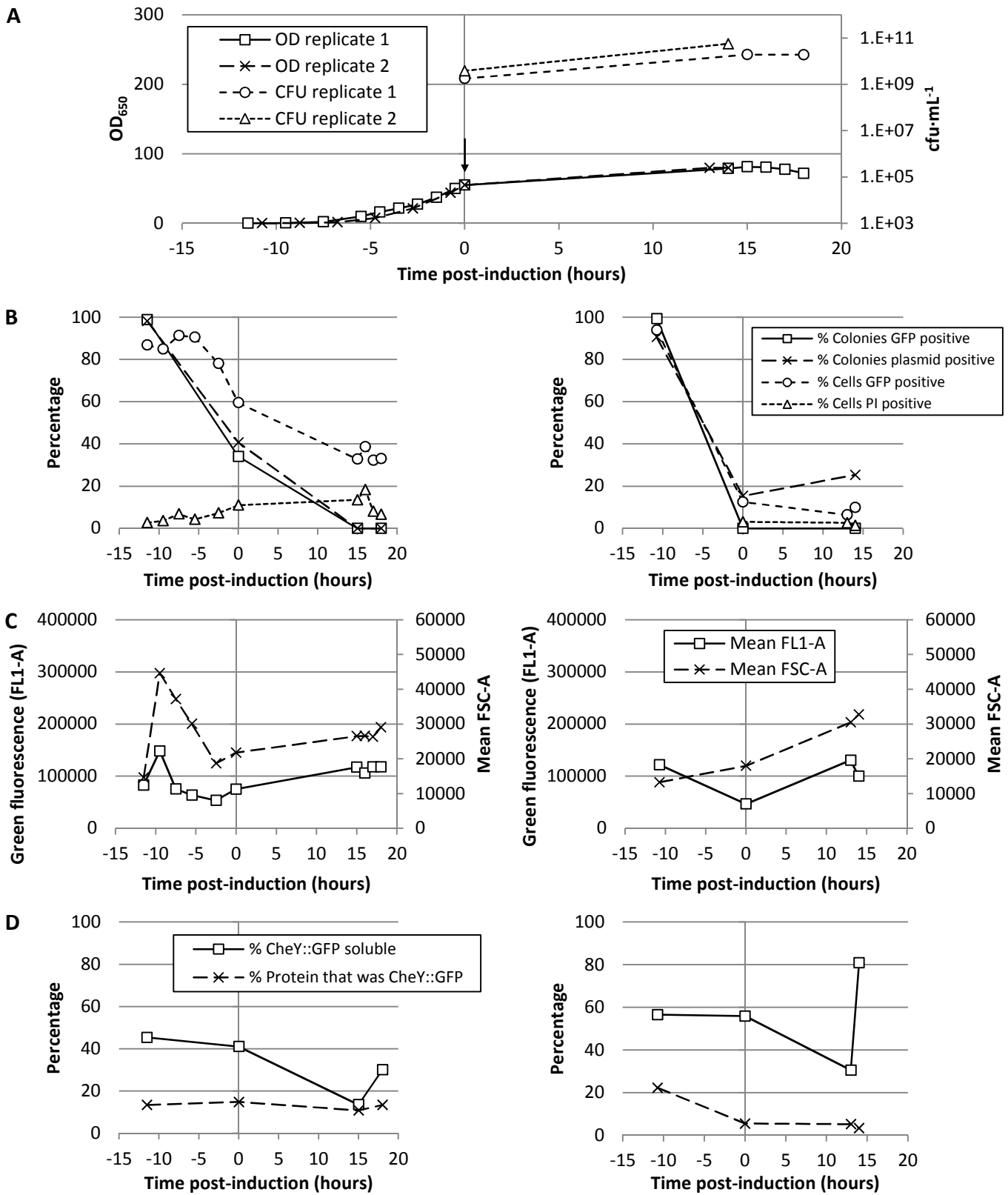[c] Estimated from DCW and % CheY::GFP of total protein assuming protein comprises 60% of *E. coli* dry cell mass (based on 50-61% estimates from [26] and 70% from [20]).
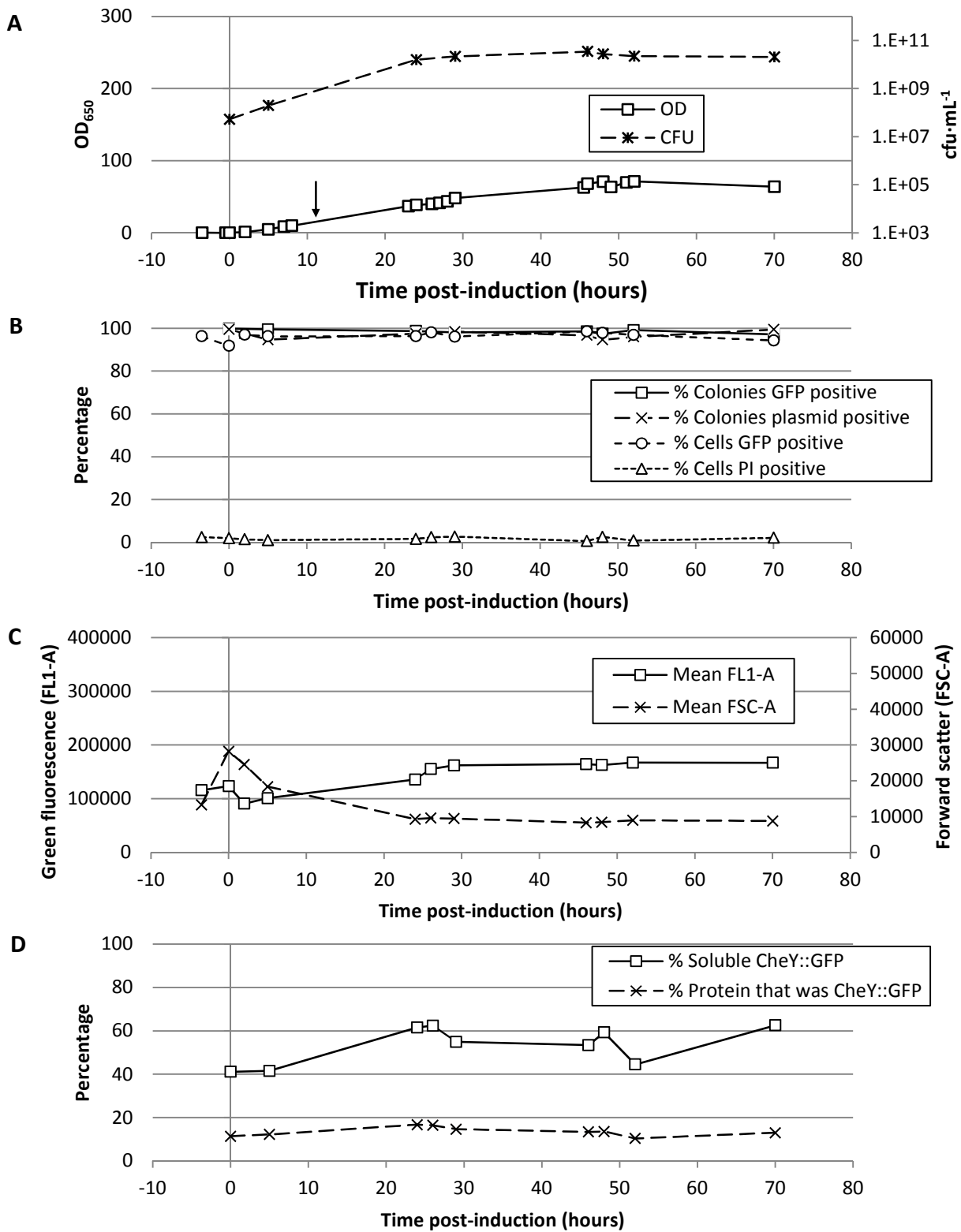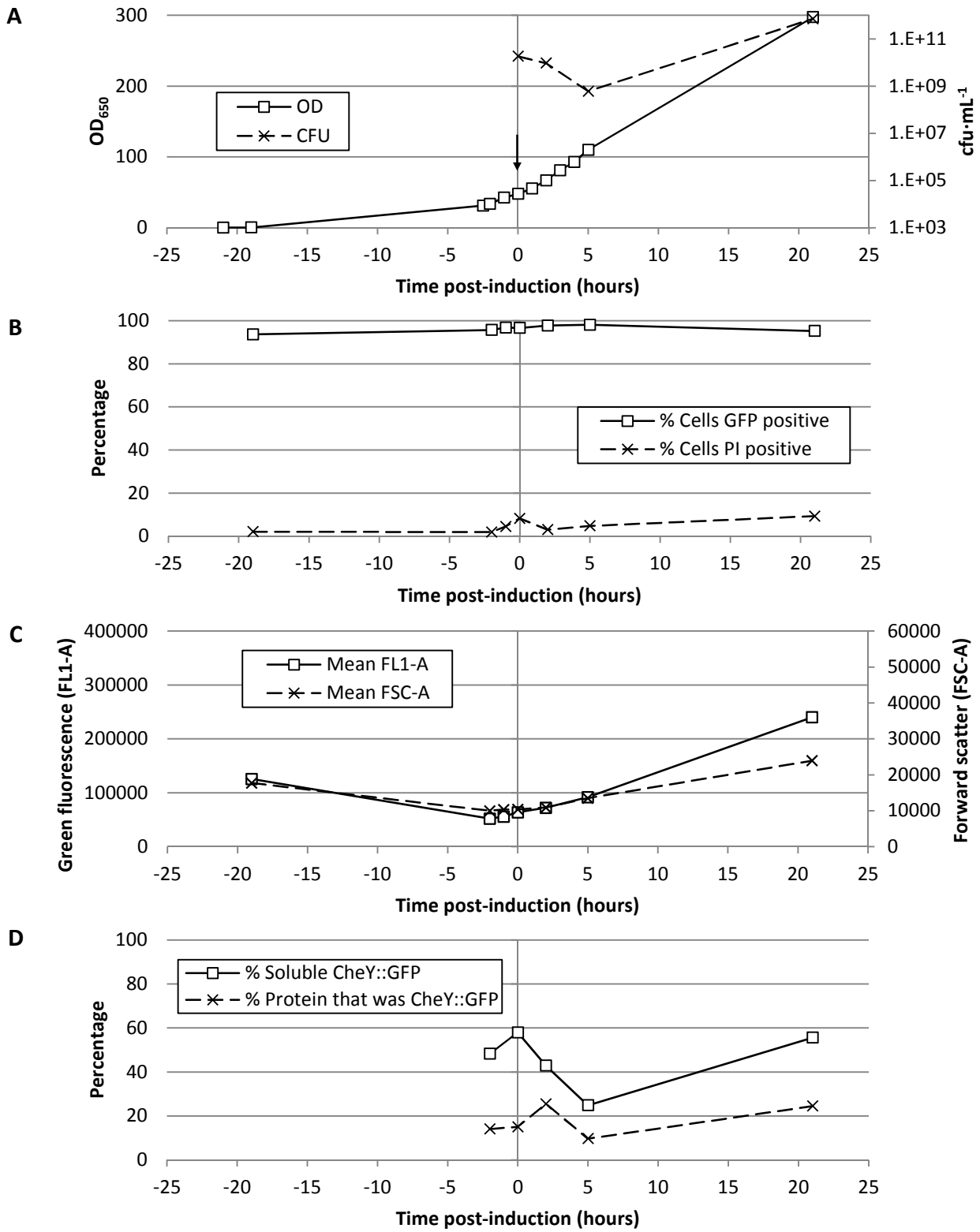
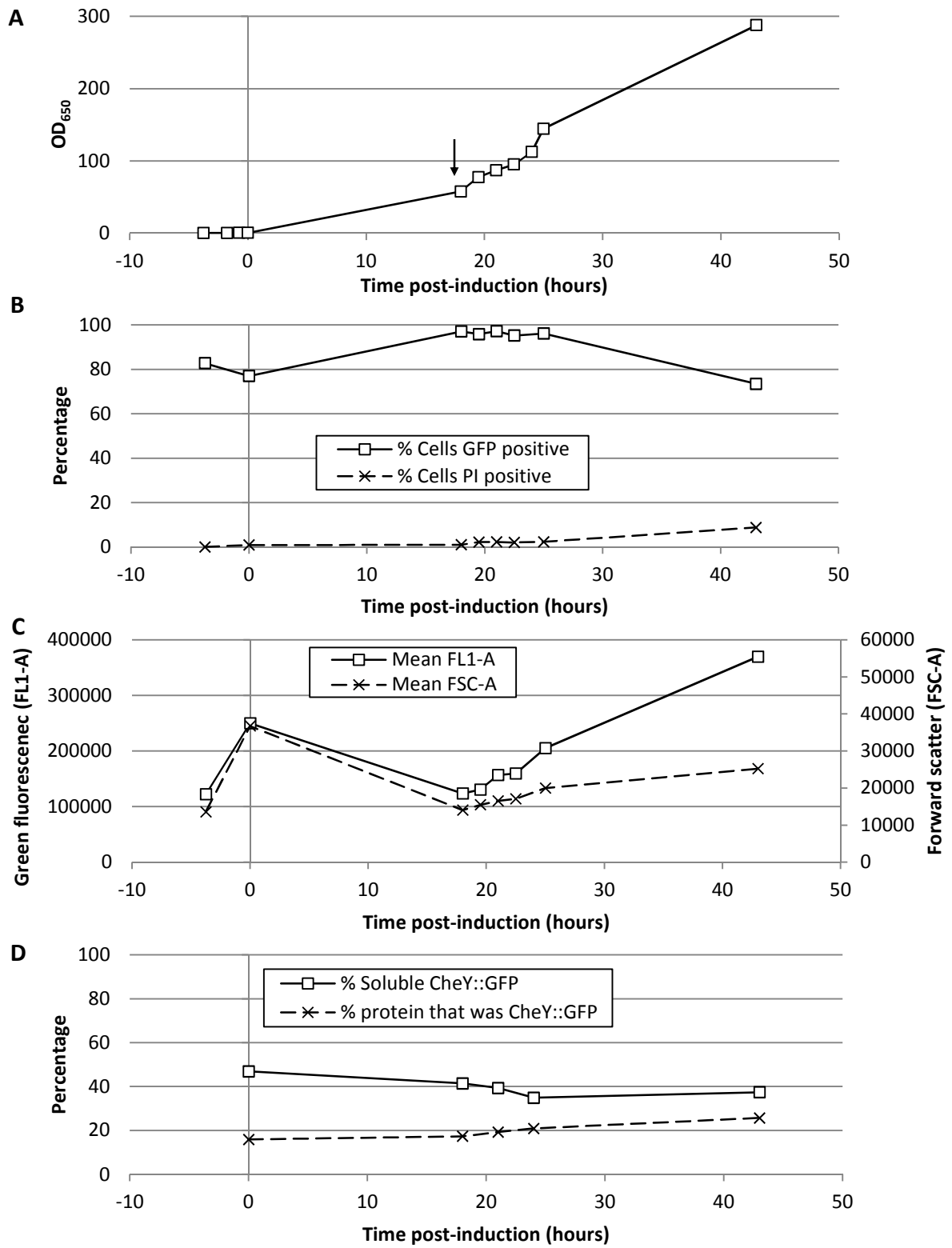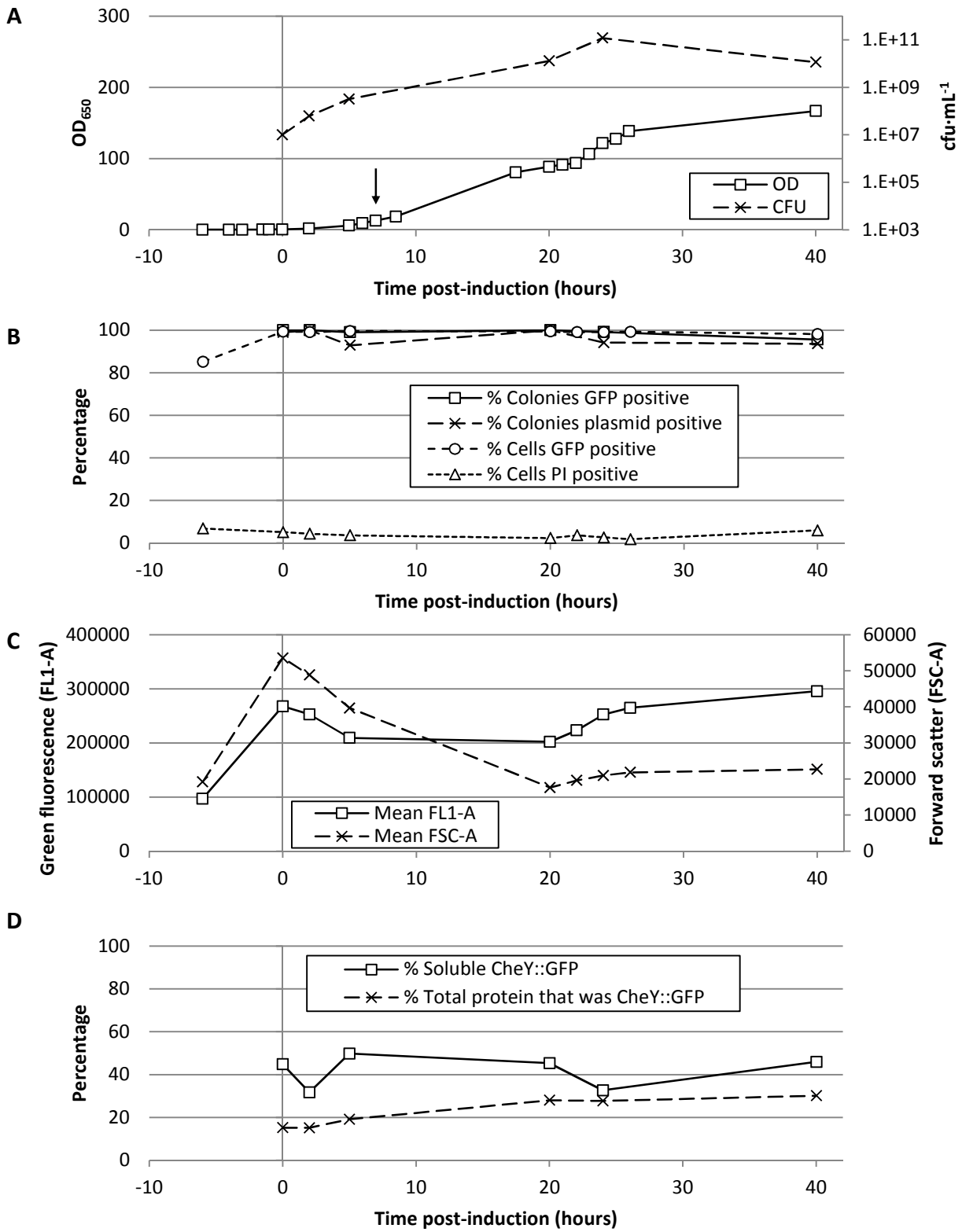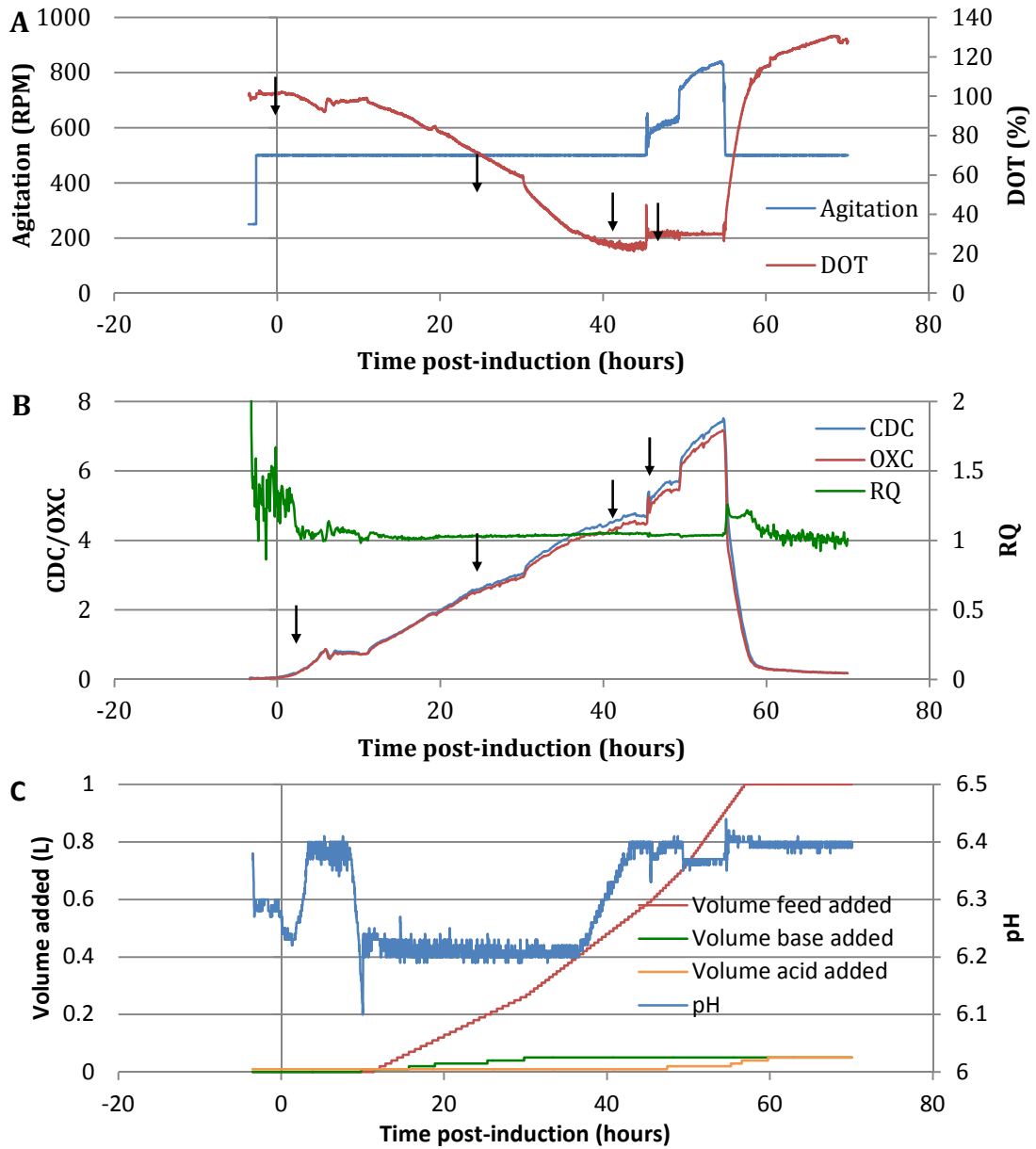Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Use of a stress-minimisation paradigm in high cell density fed-batch *Escherichia coli* fermentations to optimise recombinant protein folding

Chris Wyre and Tim W Overton*

Bioengineering, School of Chemical Engineering, and Institute of Microbiology & Infection, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

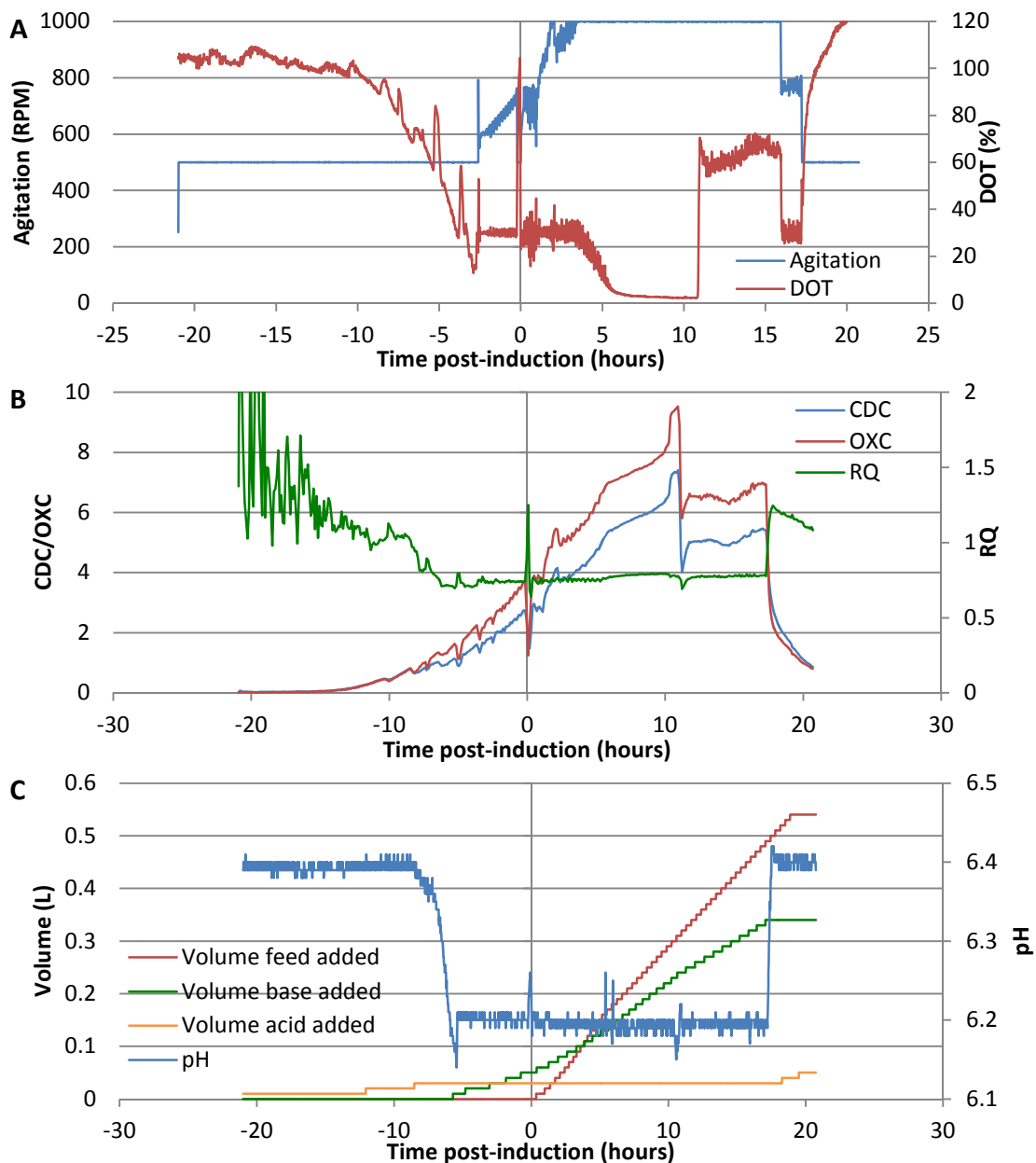*Corresponding author: t.w.overton@bham.ac.uk

Supplemental Figures 1 - 5

Supplemental figure 1. On-line data from a protocol A fermentation.
A - Agitation (blue) and DOT (red) data. B - Gas-MS data; CDC (blue) OXC (red), RQ (green). C - pH (blue), volumes of feed (red), base (green) and acid (orange) added to vessel. Arrows indicate points at which feed rate was increased, data from a representative fermentation of a minimum of 2 replicates.
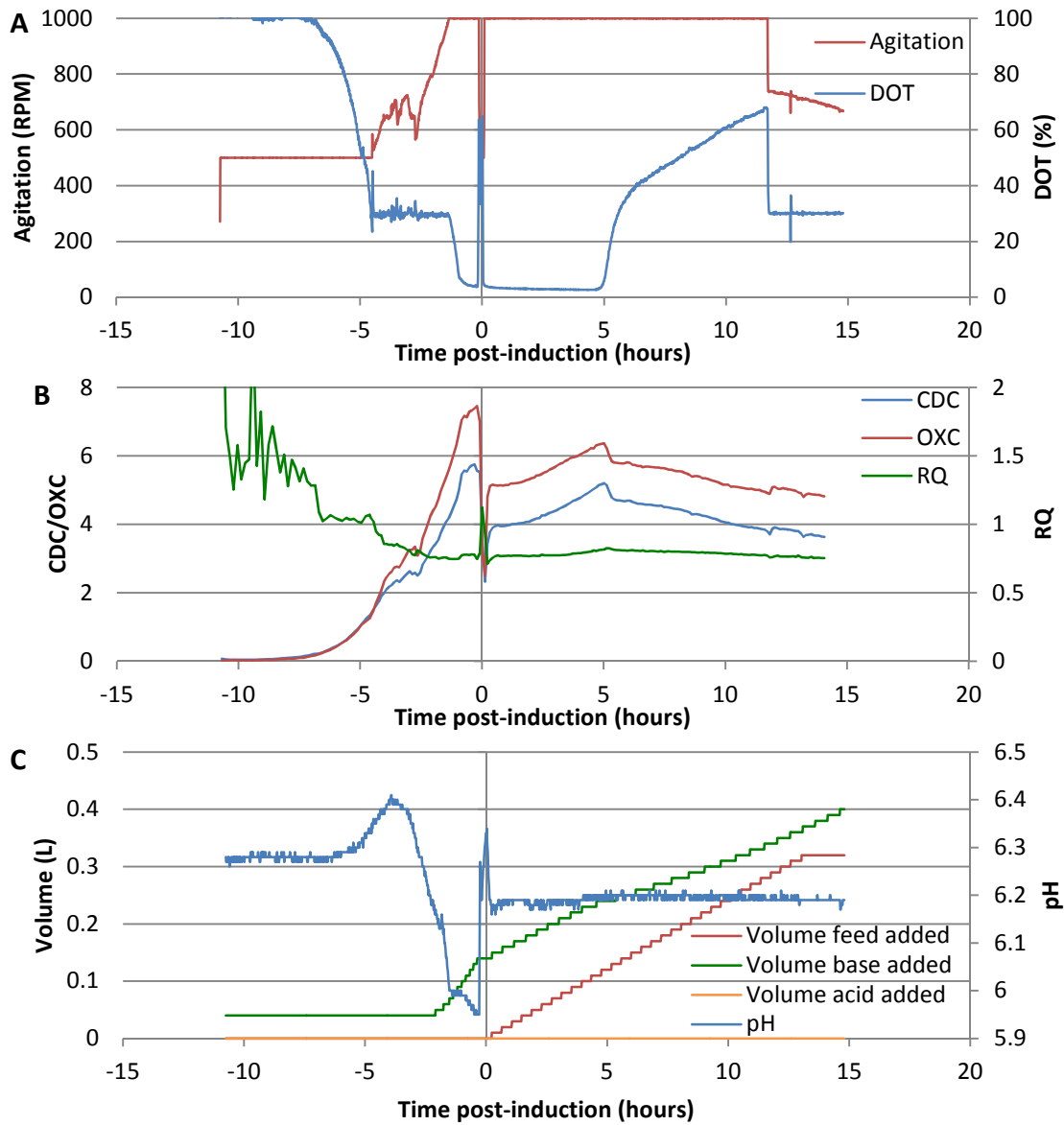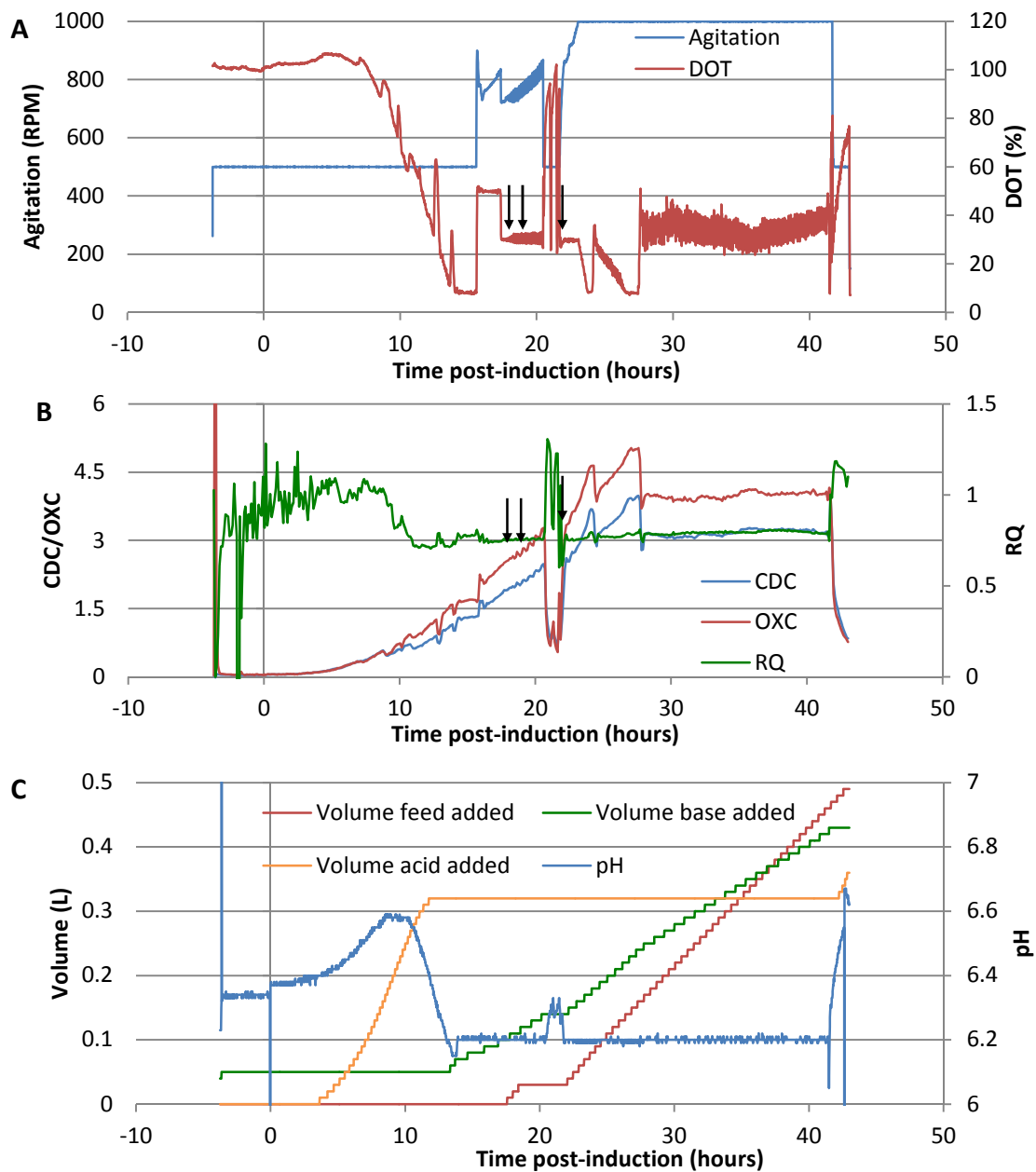
Supplemental figure 2. On-line data from a protocol B fermentation.
A - Agitation (red) and DOT (blue) data. B - Gas-MS data; CDC (blue) OXC (red), RQ (green). C - pH (blue), volumes of feed (red), acid (green) and base (orange) added to vessel.
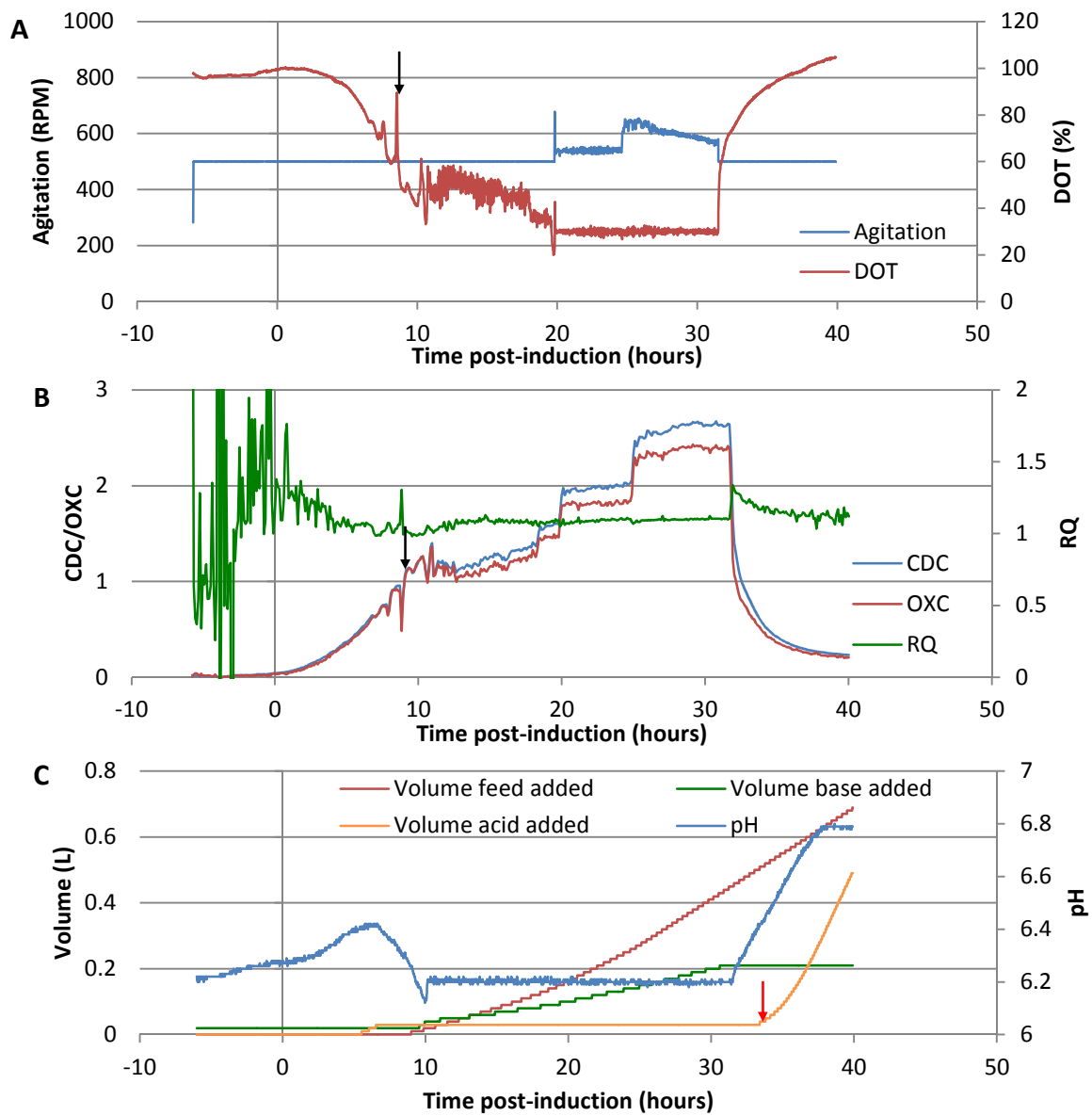
Supplemental figure 3. On-line data from a protocol C fermentation.
A - Agitation (blue) and DOT (red) data. B - Gas-MS data; CDC (blue) OXC (red), RQ (green). C - pH (blue), volumes of feed (red), base (green) and acid (orange) added to vessel. Data from a representative fermentation of a minimum of 2 replicates.

Supplemental figure 4. On-line data from a protocol D fermentation.
A - Agitation (blue) and DOT (red) data. B - Gas-MS data; CDC (blue) OXC (red), RQ (green). C - pH (blue), volumes of feed (red), acid (green) and base (orange) added to vessel. Data from a representative fermentation of a minimum of 2 replicates. Arrows indicate points at which feeding began or was paused (see text for details).

Supplemental figure 5. On-line data from a protocol E fermentation.
A - Agitation (blue) and DOT (red) data. B - Gas-MS data; CDC (blue) OXC (red), RQ
(green). C - pH (blue), volumes of feed (red), base (green) and acid (orange) added to
vessel. Black arrows indicate feeding, red arrow indicates point at which acid line
became blocked and so after which no acid could be added to vessel. Acid addition
trace is derived from fermenter software and reflects the revolutions of the acid
addition pump. Data from a representative fermentation of a minimum of 2
replicates