CamOptimus: A Tool for Exploiting Complex Adaptive Evolution to Optimise Experiments and Processes in Biotechnology

Ayca Cankorur-Cetinkaya, Joao M L Dias, Jana Kludas, Nigel K.H. Slater,Juho Rousu, Stephen G. Oliver*, Duygu Dikicioglu

Supplementary Material

Text S1. Supplementary information on test case

Data S1. Raw data for GA and 4 batches of 10 models in SR for each individual objective

This spreadsheet is comprised of 4 tabs each dedicated to the batch of 10 "evolutions" achieved for a specific objective of the combined objective function; final cell density during the growth-promoted phase (ODbi tab), further growth during the protein production phase (ODai-ODbi tab), enzyme activity (Ea tab), and specific productivity (SP tab). The models and the cognate R^2 values are provided along with the individual compositions in the "population" in 3 "generations" and the experimental values measured for each individual objective.

Movie S1. Demo video for CamOptimus

Text S1

Determination of the biological objectives for recombinant protein production under inducible promoters

The expression system employed made use of a strong methanol-inducible promoter to promote recombinant protein production and suggested that the fermentation should be divided into two main phases. The culture should first undergo a biomass-generation phase, in which the maximum possible cell density would be attained for use in the second, product-generation phase. In this second phase, metabolic flux should be directed away from cell

26 growth and towards the production of the recombinant protein once its synthesis had been
27 induced. This fermentation regime suggested that four main biological objectives should be
28 selected for our system: (i) maximising cell density prior to induction of recombinant protein
29 synthesis, (ii) maximising the biological activity of the recombinant enzyme at the end of the
30 fermentation, (iii) maximising the culture's specific productivity, and (iv) minimising any
31 further increase in cell density during the induction phase.

32

33 **Selection of factors and their boundaries to be employed in GA**

34 Published concentrations of medium components that acted as the major nitrogen, potassium,
35 magnesium, and calcium sources in *K. phaffii* varied substantially in studies employing
36 different clones or different growth strategies [1–4]. Among the micronutrients, iron
37 displayed the largest variation in concentration between different defined medium recipes
38 [4,5]. Cells that employ the *AOX* promoter to express the recombinant protein have usually
39 been cultivated on sorbitol or glycerol prior to induction, with a co-feeding strategy based on
40 methanol/sorbitol mixtures employed during the induction period [3, 6]. Preliminary shake-
41 flask experiments revealed that HuLy activity was lost from batch cultures whose pH was not
42 controlled throughout the course of the fermentation, and so pH was included as another
43 factor to be optimised.

44 The maximal and the minimal concentrations of medium components previously reported in
45 the literature were used to set the allowable ranges for each factor [4,5,7–9] Although pH
46 values between 5 and 6 have been used in most studies, *K. phaffii* can grow in a wide range
47 pH from 3 to 7 [10]. The pH of the culture can affect proteolytic activity, secretion, and
48 protein production [11]; therefore, we kept the pH range as wide as possible in order to
49 identify the optimum value for HuLy production.

50

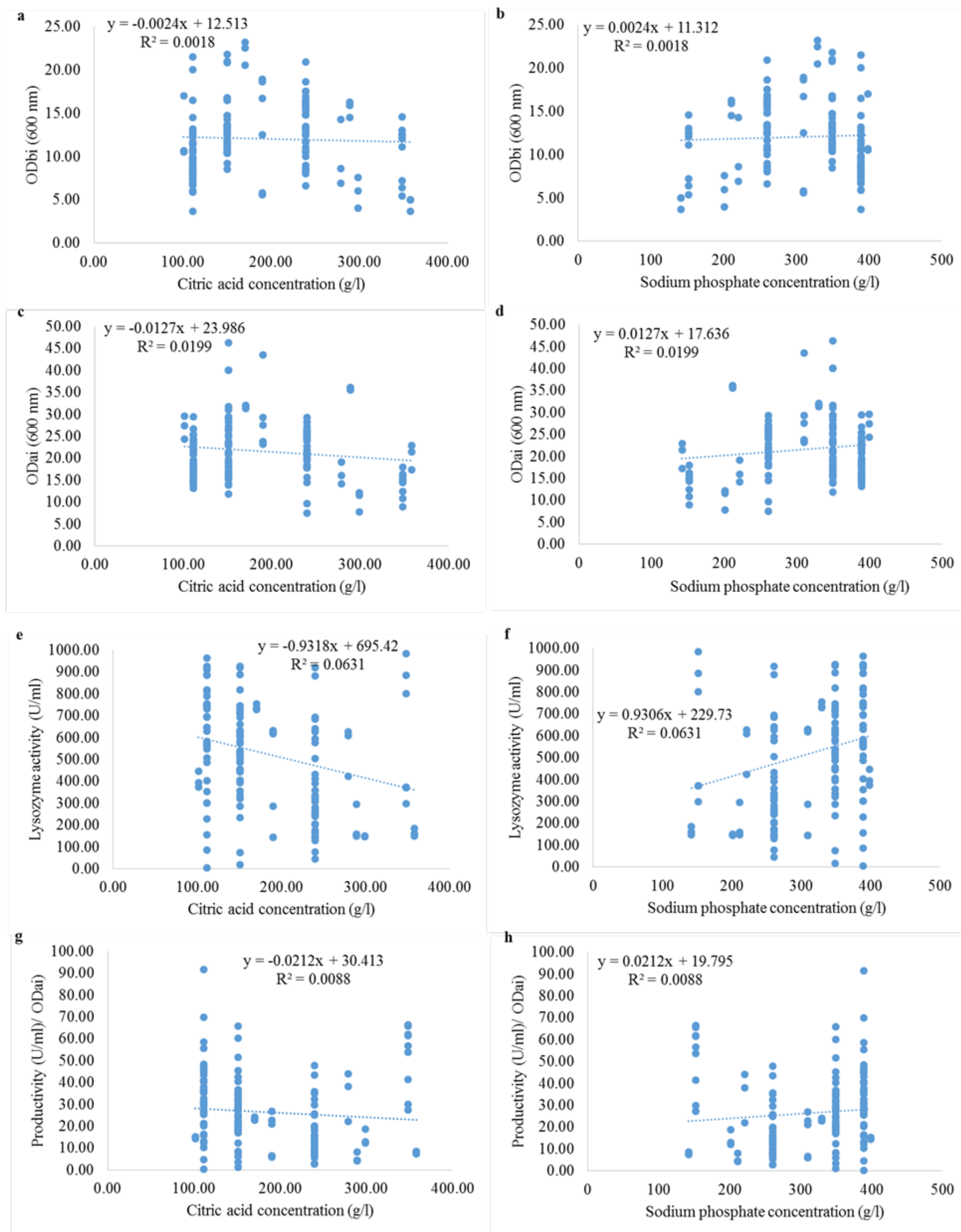
Investigation of the effect of the citrate/phosphate buffer on culture performance

The effect of the addition of buffer components to maintain constant pH on the growth and production characteristics of the culture was investigated at different stages of the analysis. Preliminary analysis involved growing the microbial culture in shake flasks, where the culture pH was maintained constant in buffered conditions, or was allowed to decrease in non-buffered defined medium. The protein activity diminished in cultures where the pH was not kept constant, thus demonstrating the necessity for pH control. For the operating conditions investigated in this study, the working pH was in the range of 3-7. Citrate/phosphate buffer, with reported functionality in the pH range of 2.6-7 [12], was selected as a suitable option for our purposes, since its constituents did not pose a threat of toxicity for the culture. Since citric acid and dibasic sodium phosphate were used to prepare the buffer, possible interference due to buffer components acting as macronutrients was investigated to rule out such complications. Varying the concentration of citric acid (5-fold) to account for the possible variation it would display in the pH range of 2.6 – 7 (<5-fold) did not yield a change in growth or protein activity. In order to investigate the effect of the phosphate group acting as a macronutrient, a similar experiment was carried out as described for testing the effect of the citrate component of the buffer. Furthermore, in order to test the phosphate group specificity of the analysis, dibasic sodium phosphate was replaced by sodium sulphate, to provide equivalent Na molarities. The difference in growth or protein activity always remained within 10% of one another and that of the reference study, regardless of the concentrations employed within the limits that were of interest for the purpose of this study (Table A). Therefore the citrate/phosphate buffer was considered as an acceptable option for the cultivations performed in snap-cap tubes. Throughout the evolutionary procedure, the pH of randomly selected cultures was measured in order to ensure that the buffer was working satisfactorily. The complete data collected over three

generations were investigated for the possible existence of any correlation between any single objective and the buffer components individually and no trends were detected with all regression coefficients (R^2) remaining <0.10 (Figure A). We then extended our investigation of the citrate component further by tracing any possible indication that its chelating properties impacted on the availability of calcium or iron, both of which were under consideration within the scope of this work. Despite earlier reports on the calcium- and iron-chelating properties of citric acid [13], the concentrations employed in the study were shown not to cause such an effect throughout the course of the study, with the Pearson correlation coefficients for both cases remaining below 0.3 and the regression coefficients below 0.1 (Figure B).

Table A. Effect of buffer components on growth and protein production.

	% change in OD	**% change in HuLy activity**
Citric acid supplementation	2.362	8.456
NaH_2PO_4 supplementation	5.133	7.730
Na_2SO_4 supplementation	7.605	4.351



89

90 **Figure A. Investigation of the possible correlation between individual objectives and the**

91 **concentration of the buffer components.**

92 Complete data set comprised of all measured objectives over three generations. The changes

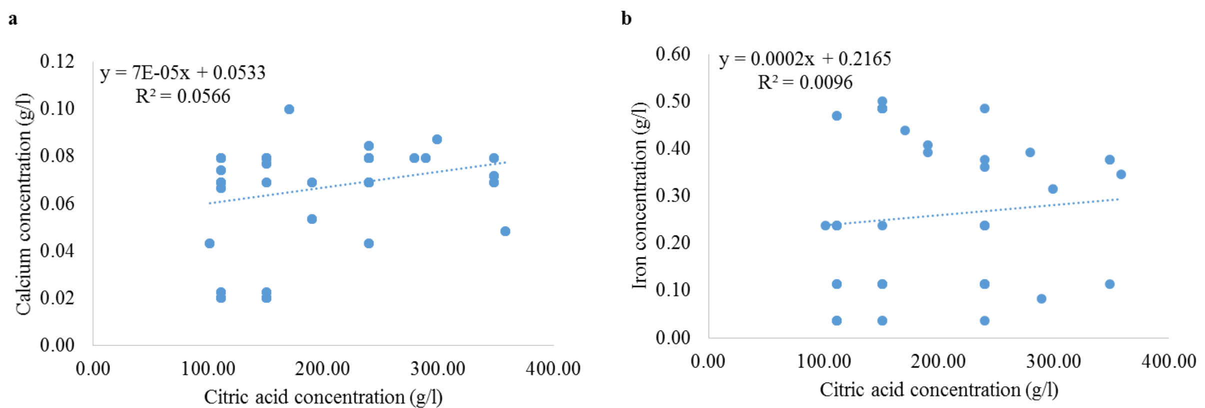
93 in biomass during the growth promoted phase (OD_{bi}) as a function of citric acid concentration

94 (a) or dibasic sodium phosphate concentration (b), the final biomass attained (OD_{ai}) as a

95 function of citric acid concentration (c) or dibasic sodium phosphate concentration (d), the

96 HuLy activity as a function of citric acid concentration (e) or dibasic sodium phosphate

97 concentration (f), and the culture productivity as a function of citric acid concentration (g) or
98 dibasic sodium phosphate concentration (h) are provided along with their linear regression
99 models and R^2 is provided as a goodness-of-fit measure.



100

101 **Figure B. Investigation of the possible chelation effect of citric acid on calcium and iron**
102 **in the cultivation medium.**
103 Distribution of calcium (a) and iron (b) concentrations tested over three generations as a
104 function of the citric acid concentrations of the corresponding sample. Any possible trends
105 and correlations were investigated by fitting linear models (R^2 as a goodness-of-fit measure).
106 The slopes of the models indicating the existence of possible trends were negligibly small.

107

108 **Fine-tuning the global optimum by population profiling**

109 We carried out fine-tuning of the optimised environmental conditions that we had obtained at
110 the end of the 3rd "generation" in the GA study by employing a population profiling
111 methodology to determine whether the performance of the system could be improved even
112 further. We investigated the change in how many individuals assumed each level over the 3
113 "generations" for each one of the 9 factors (Figure 2). We calculated the percent occupancy
114 of the levels in the better-performing, i.e. the "fitter" fraction of the "population" in the last
115 "generation", which corresponded to the half of the "population" with the highest "fitness"

116 scores, and employed this as a footprint for identifying the optimal levels for each factor
117 (Figure C). The footprints of methanol, sorbitol and pH in the 3rd "generation", as well as
118 their convergence profiles through the course of the GA search, indicated that the levels that
119 these factors assumed converged towards unique values at 6.75g/L, 7.60g/L and 6.74,
120 respectively (Figure 2(a-c) and Figure C(a-c)). In the case of ammonium, potassium and
121 glycerol, although the convergence profiles displayed bi-modal behaviour, the footprints in
122 the 3rd "generation" indicated a pronounced convergence towards unique levels at 6.55 g/L,
123 3.21 g/L and 9.87 g/L, respectively (Figure 2(d-f) and Figure C(d-f)). The convergence of the
124 footprint distribution for the remaining 3 factors either were not as pronounced, e.g. iron
125 (Figure C(g)), or displayed bi-modal behaviour, e.g. calcium and magnesium (Figure C(h-i)).
126 The optimal concentration of FeSO4.7(H_2O) was determined as 0.04 g/L. The distribution of
127 the frequency of the level occupation over successive generations was investigated for the
128 factors, whose footprints displayed bi-modal distributions. The level of 0.07g/L for
129 $CaCl_2$.2(H_2O) was employed more frequently in the "population" as the search approached
130 the 3rd generation (Figure C(h)) and therefore this level was selected as the optimal
131 concentration for $CaCl_2$.2(H_2O). A similar inspection of the progression of the population
132 profiles for $MgSO_4$.7(H_2O) indicated that two distinct levels were more frequently adopted as
133 the system evolved towards an optimum (Figure C(i)). Therefore, both levels were tested
134 experimentally to identify the better performing condition. The lower of these two levels
135 improved the activity of the secreted HuLy enzyme and the specific productivity by nearly
136 20- and 15-fold, respectively. The adoption of the lower value also resolved issues related to
137 the medium precipitating, thus substantially improving reproducibility. The $MgSO_4 7 \cdot H_2O$
138 level for the optimal "individual" was selected as 2.37g/L.

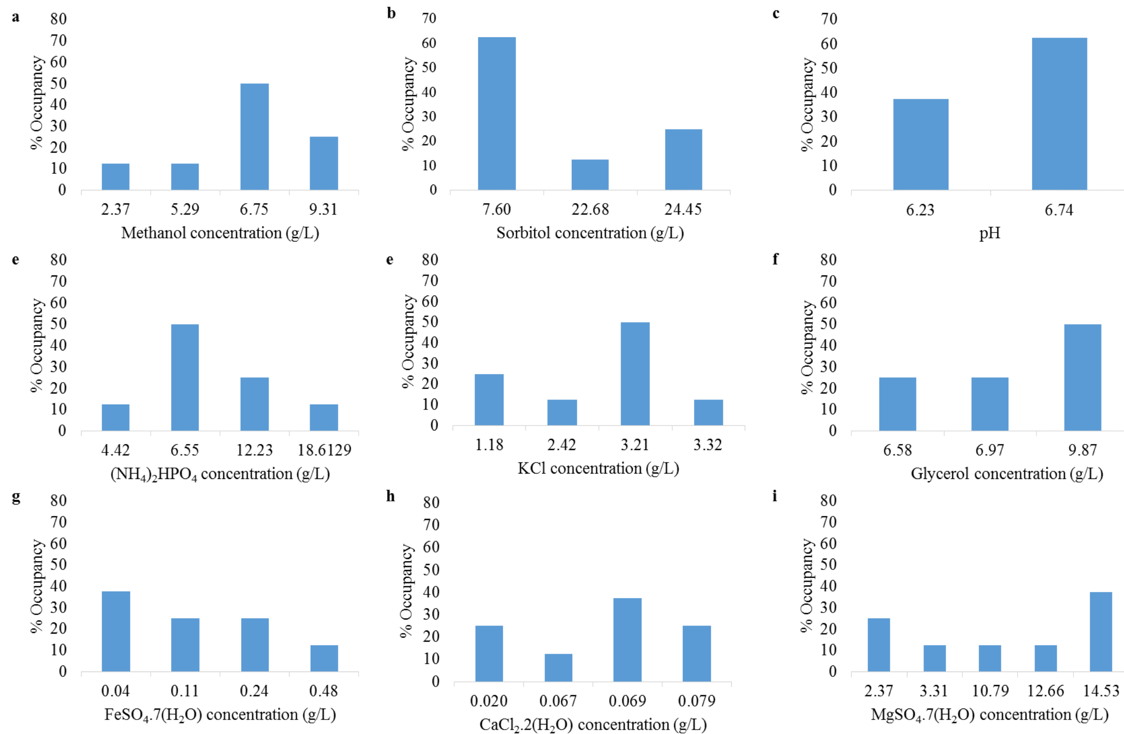


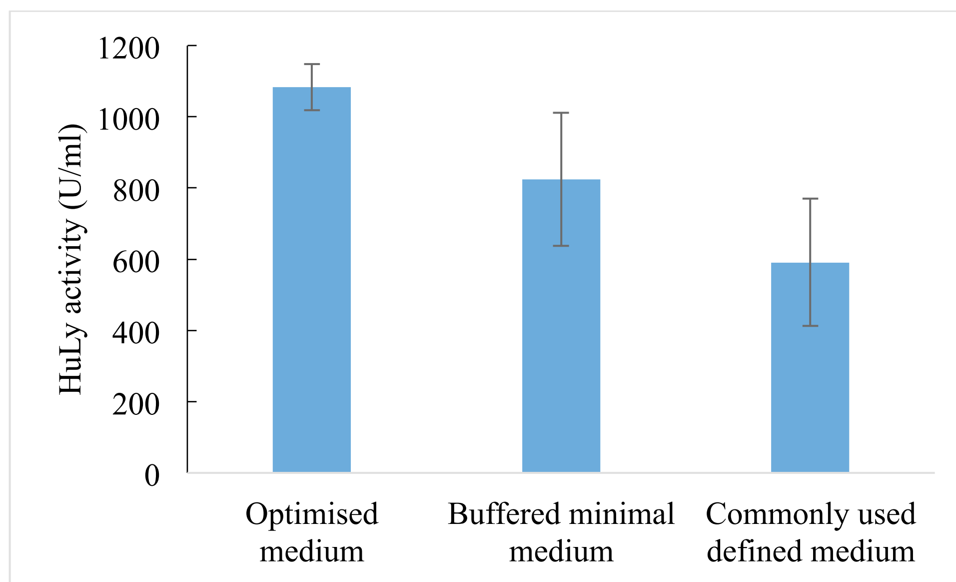
Figure C. Relative occupancy of the levels of each factor in the better performing fraction of the 3rd generation

Relative occupancy of the levels assigned in the better performing fraction are displayed (i.e. the 50% of the population with the highest "fitness" scores, of the 3rd generation for each factor). Methanol (a), sorbitol (b), pH (c), $(NH_4)_2PO_4$ (d), KCl (e), glycerol (f), $FeSO_4.7(H_2O)$ (g), $CaCl_2.2(H_2O)$ (h), and $MgSO_4.7(H_2O)$ (i). The footprint was generated by plotting the relative distribution of the instances a given level was employed in the better-performing half of the 3rd generation. The levels, which appeared at least once in the last generation, is displayed on the abscissa and the % occupancy, indicating the relative representation of each level, is displayed in the ordinate.

Challenging the optimised cultivation conditions against the existing conditions

Buffered minimal medium, containing yeast nitrogen base, and defined medium [4,14] were frequently reported as the cultivation environment for *K. phaffi* strains expressing

153 recombinant proteins under the control of the AOX promoter. pH values between 5 and 6 are
154 used in most studies involving *K. phaffii* fermentations [10]. We challenged the optimised
155 environmental conditions against these commonly employed conditions to assess the
156 performance of the methodology. We showed that the activity of the secreted enzyme in the
157 culture could be increased by 30% and 80% through adopting the optimised conditions to
158 replace minimal medium buffered at pH 6 and defined medium buffered at pH 5, respectively
159 (Figure D).



160

161 **Figure D. Challenging the performance of *K. phaffii* cultivated under optimised**
162 **conditions.**

163 The error bars represent the variation between the technical triplicates. The variation in the
164 cultures investigated under the optimised conditions was the lowest among the three different
165 cultivation conditions that were tested.

166 **Determination of Population Size and Number of Generations for SR Analyses**

167 In order to identify the suitable parameter setting, regression analyses were conducted for a
168 number of combinations of population size and generation number to model the protein

169 activity (Table B). The goodness of fit values were provided as an average of 3 independent
170 simulations for each setting.

171 **Table B.** Effect of Population Size and Number of Generations on Goodness of Fit

Population Size	Number of Generations	Goodness of Fit of the Model
100	100	0.772±0.022
100	500	0.799±0.007
500	100	0.811±0.018
500	500	0.814±0.012
500	1000	0.795±0.025
1000	500	0.796±0.032
1000	1000	0.810±0.014

172

173 **Sensitivity of each objective to each factor as determined by SR models**

174 Growth yield during the protein production phase was determined to be highly sensitive to
175 changes in the concentrations of glycerol and ammonium in the cultivation environment. This
176 was an expected outcome since both the preferred carbon source (glycerol) and the preferred
177 nitrogen source (ammonium) become limiting during that phase in the fed-batch culture
178 leading to incremental changes that are readily reflected in the growth yield. Enzyme activity
179 was also highly sensitive to the variations in the availability of glycerol and ammonium since
180 the production of the recombinant protein necessitated competing with growth for the
181 available resources. In addition, enzyme activity was also highly sensitive to variations in
182 methanol concentration, which was also expected since the production system was activated
183 by methanol inducing the alcohol oxidase promoter. Specific productivity, a compound
184 objective, was highly sensitive to variations in all potential carbon sources; both available

185 (methanol and sorbitol), and unavailable (glycerol) during the protein production phase.
186 Additionally, we identified specific productivity to be highly sensitive to variations in the
187 concentration of calcium. This was an unexpected outcome since Human Lysozyme has been
188 reported not to bind calcium [15]. Moreover, calcium was reported to merely contribute to the
189 stability of calcium-binding lysozymes [16]. The results highlighted here may suggest the
190 existence of a possible universal role for calcium for the lysozyme family, which requires
191 further investigation. The models constructed in this study were not able to capture the
192 variance in the objectives that was caused by potassium, iron, or magnesium since none of
193 the objectives were identified to be sensitive to variations in either one of these predictors.
194 Specifically, magnesium was not included as a predictor in many of the constructed models
195 (Figure E) and the bi-modal distribution of its levels in the search-space of the GA was
196 thought to contribute to that effect.
197

		Factor	model 1	model 2	model 3	model 4	model 5	model 6	model 7	model 8	model 9	model 10
a	OD bi	Ammonium	0.89	0.29	1.28	0.85	0.00	1.53	1.64	0.32	1.16	0.09
		Potassium	1.56	0.66	3.10	2.45	2.98	2.10	0.43	2.70	0.16	5.61
		Magnesium	1.04	0.14	1.01	1.89	0.17	0.51	0.26	0.84	N/A	0.05
		Iron	3.15	0.05	0.01	0.04	2.77	4.25	2.54	0.08	1.61	0.06
		Calcium	1.89	1.02	1.66	1.77	2.07	2.90	N/A	0.08	0.57	1.69
		Glycerol	0.03	3.56	1.82	0.40	9.85	6.14	N/A	0.95	9.57	0.20
		pH	31.30	61.03	98.32	29.67	23.10	48.18	52.06	12.71	96.50	94.58
b	OD ai-bi	Ammonium	6.37	12.62	N/A	0.19	0.57	0.45	N/A	0.03	1.45	1.77
		Potassium	1.54	0.64	0.95	0.07	4.56	0.45	3.84	0.19	1.68	3.98
		Magnesium	0.02	0.38	0.99	1.71	0.15	2.08	0.47	0.55	0.06	1.36
		Iron	3.19	0.15	0.13	0.04	3.94	0.03	0.24	0.13	0.26	0.55
		Calcium	N/A	0.66	1.77	1.42	0.80	3.04	4.06	0.43	4.54	2.15
		Glycerol	1.69	41.97	8.72	5.28	4.32	0.70	15.27	6.08	15.00	21.16
		Methanol	3.71	2.13	0.11	2.61	3.10	4.87	N/A	0.08	N/A	N/A
		Sorbitol	0.11	1.43	0.56	3.40	1.12	3.30	0.06	1.66	0.66	3.97
		pH	0.49	15.67	14.58	0.08	5.01	5.34	0.92	2.19	5.08	11.13
c	Enzyme activity	Ammonium	1.06	0.96	5.68	16.19	4.65	0.20	1.30	0.20	1.48	1.07
		Potassium	0.70	0.06	2.81	0.76	1.27	0.98	0.76	0.95	0.87	0.51
		Magnesium	N/A	N/A	0.33	N/A	0.30	0.19	0.10	0.09	N/A	0.65
		Iron	2.65	0.04	0.08	2.40	1.58	3.89	1.01	2.19	4.00	1.47
		Calcium	3.63	0.78	0.93	2.64	7.07	0.38	2.58	2.14	0.96	0.16
		Glycerol	9.63	10.77	8.22	0.92	6.96	18.13	11.97	8.80	20.90	5.63
		Methanol	29.24	N/A	5.22	3.14	2.97	9.98	0.31	0.13	13.17	4.39
		Sorbitol	1.05	2.90	1.17	1.10	0.14	6.20	1.35	0.47	1.33	0.02
		pH	28.12	49.58	6.77	6.75	11.33	24.25	8.17	29.43	26.87	2.78
d	Specific productivity	Ammonium	5.11	1.28	0.95	1.58	3.96	1.21	0.66	5.79	0.99	2.16
		Potassium	1.87	4.51	0.29	1.70	0.41	1.63	0.01	0.20	4.22	0.13
		Magnesium	0.16	N/A	N/A	N/A	N/A	N/A	0.02	0.55	0.08	0.10
		Iron	0.01	0.15	3.01	2.76	2.31	0.37	0.57	0.05	4.20	3.44
		Calcium	N/A	0.25	1.39	0.03	0.67	0.20	N/A	0.30	10.61	1.26
		Glycerol	7.12	17.22	9.72	3.13	2.59	11.69	4.19	6.35	22.04	0.09
		Methanol	2.73	7.99	35.51	3.55	3.07	2.94	61.94	60.90	5.34	7.63
		Sorbitol	85.25	2.02	2.79	3.45	0.15	2.29	0.01	4.82	23.42	2.99
		pH	3.69	11.66	29.83	2.70	8.53	8.37	60.81	64.46	7.79	N/A

Figure E. Heat map of the sensitivity of each individual objective to variations in each individual factor in the model pools

The sensitivity of each objective [final cell density during the growth-promoted phase (ODbi) (a), further growth during the protein production phase (ODai-ODbi) (b), enzyme activity (Ea) (c) and specific productivity (SP) (d)] to a variation in the value of each factor of ±10%. The shade of green denotes the strength of the response, the corresponding values numerically denote its strength, in percent terms, for each model.

References

1. **Cos O, Ramón R, Montesinos JL, Valero F**. *Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast Pichia pastoris*

210 *under different promoters: a review. Microb Cell Fact*. 2006;5:17.

211 2. **Gach JS, Maurer M, Hahn R, Gasser B, Mattanovich D, et al.** *High level*
212 *expression of a promising anti-idiotypic antibody fragment vaccine against HIV-1 in*
213 *Pichia pastoris*. *J Biotechnol*. 2007;128(4):735–46.

214 3. **Hesketh AR, Castrillo JI, Sawyer T, Archer DB, Oliver SG**. *Investigating the*
215 *physiological response of Pichia (Komagataella) pastoris GS115 to the heterologous*
216 *expression of misfolded proteins using chemostat cultures*. *Appl Microbiol Biotechnol*.
217 2013;97(22):9747–62.

218 4. **Prielhofer R, Maurer M, Klein J, Wenger J, Kiziak C, et al.** *Induction without*
219 *methanol: novel regulated promoters enable high-level expression in Pichia pastoris.*
220 *Microb Cell Fact*. 2013;12(1):5.

221 5. **Delic M, Mattanovich D, Gasser B**. *Monitoring intracellular redox conditions in the*
222 *endoplasmic reticulum of living yeasts. FEMS Microbiol Lett*. 2010;306(1):61–6.

223 6. **Niu H, Jost L, Pirlot N, Sassi H, Daukandt M, et al.** *A quantitative study of*
224 *methanol/sorbitol co-feeding process of a Pichia pastoris Mut+/pAOX1-lacZ strain.*
225 *Microb Cell Fact*. 2013;12(1):33.

226 7. **Baumann K, Maurer M, Dragosits M, Cos O, Ferrer P, Mattanovich D**. *Hypoxic*
227 *fed-batch cultivation of Pichia pastoris increases specific and volumetric productivity*
228 *of recombinant proteins*. *Biotechnol Bioeng*. 2008;100(1):177–83.

229 8. **Pfeffer M, Maurer M, Stadlmann J, Grass J, Delic M, et al.** *Intracellular*
230 *interactome of secreted antibody Fab fragment in Pichia pastoris reveals its routes of*
231 *secretion and degradation*. *Appl Microbiol Biotechnol*. 2012;93(6):2503–12.

232 9. **Çelik E, Çalik P, Oliver SG**. *Fed-batch methanol feeding strategy for recombinant*
233 *protein production by Pichia pastoris in the presence of co-substrate sorbitol*. *Yeast*.

234 2009;

235 10. **Cregg JM, Vedvick TS, Raschke WC**. *Recent Advances in the Expression of Foreign*
236 *Genes in Pichia pastoris*. *Bio/Technology*. 1993;11(8):905–10.

237 11. **Fakruddin M, Mazumdar MR, Mannan BKS, Chowdhury A, Hossain MN**.
238 *Critical Factors Affecting the Success of Cloning, Expression, and Mass Production of*
239 *Enzymes by Recombinant E. coli*. *ISRN Biotechnol*. 2013;(3):1–7.

240 12. **Pearse A**. *Histochemistry: Theoretical and applied. Am J Med Sci*. 1980;97–158.

241 13. **Ash M, Ash I**. *Handbook of Green Chemicals*. *Synapse Info Resources*; 2004. 1348 p.

242 14. **Kumita JR, Johnson RJK, Alcocer MJC, Dumoulin M, Holmqvist F, et al.** *Impact*
243 *of the native-state stability of human lysozyme variants on protein secretion by Pichia*
244 *pastoris*. *FEBS J*. 2006;273:711–20.

245 15. **Kissinger CR, Parge HE, Knighton DR, Lewis CT, Pelletier LA, et al.** *Crystal*
246 *structures of human calcineurin and the human FKBP12-FK506-calcineurin complex.*
247 *Nature*. 1995;378(6557):641–4.

248 16. **Vogel HJ**. *Calcium-Binding Protein Protocols*. Vol. 172. *Humana Press*; 2002.

249