Supplementary Information

Supplementary Text S1- Comparison between the turnover dependent flux constraints used by TDPS and other algorithms that simulate up/down-regulations

With the exception of gene knockouts, which can be simulated by constraining a flux to zero, simulating the impact of other genetic modifications can be difficult when translating a genotype into flux changes. Two different general strategies has been described for simulating the effect of gene up/down-regulation: constraining the regulated flux to be above or below a certain threshold (as used by OptReg [1] and the under/over-expression plugin in OptFlux [2]) or adding the regulated fluxes to the objective function with positive or negative coefficients (as used by Redirector [3]).

By restricting the regulated fluxes to rigid values, one can infer how the rest of the network would have to adapt to a certain change. As shown in Figure S1b and S1c, up-regulating reaction R3 or R5 with OptReg results in a reduction in biomass accumulation and the excretion of metabolite T or X. However, as shown in Figure S1c, up-regulating a terminal reaction from a pathway can result in fluxes increasing in other reactions, even reactions inactive in a reference state (reaction R4 increases from 0 to 50). Although the results obtained with this type of strategy can still identify correctly genetic targets for promoting the excretion of a certain metabolite, the biological interpretation of flux distributions obtained with rigid flux constraints is much more limited. Using the case depicted in Figure S1a and S1c, one can see that in an organism that has zero availability of metabolite D, up-regulating reaction R5 should not be enough to lead to the production of metabolite X. Therefore, if the goal is to estimate the impact of a certain up-regulation, forcing fluxes to be above a certain threshold will lead to results with little biological meaning.

Another strategy to simulate up and down-regulations consists of adding the regulated reactions to the objective function with positive (up-regulation) or negative coefficients (down-regulation) as is done by Redirector. Figure S1d and S1e show possible consequences of up-regulating

reactions based on manipulating the objective function. As shown in Figure S1d, the up-regulation of reaction R3 by adding it to the objective function with half the weight of biomass results in both biomass and metabolite T production. However, applying the same protocol to the up-regulation of R5 results in the exclusive production of metabolite X. As discussed above for the up-regulation of R5 with OptReg (Figure S1c), using Redirector for up-regulating R5 also results in the activation of additional reactions up-stream of R5 (Figure S1e), which again shows the limitation of these algorithms for predicting the effect of up-regulations.

The under/over-expression plugin from OptFlux also simulates up/down-regulations by applying rigid flux constraints. However, unlike the other two algorithms described above, it cannot simulate the regulation of reactions with no flux in a reference state, which reduces its applicability in the variety of strains that can be tested. As shown in Figure S1f and S1g, both up-regulations used as examples result in no flux changes in comparison to the reference.

In order perform up/down-regulations in a way that better mimics the intracellular environment, TDPS makes the flux in regulated reactions dependent on the production fluxes of the precursor for the regulated reactions. As shown in Figure S1h, making the flux in an up-regulated reaction R3 dependent on the “availability” of metabolite C (produced by reaction R2), it is possible to redirect the consumption of this metabolite to R3 without forcing any other changes in the network. Furthermore, the flux in reaction R3 becomes limited by the production turnover of metabolite C and the only way to increase it above 80 flux units would be to add more regulations to the network to increase the production flux for this metabolite. By making regulated fluxes dependent on the “availability” of precursor metabolites, when reactions without available precursors are up-regulated, no flux is forced though them. As shown in Figure S1i, up-regulating R5 has no impact on the flux distribution, unlike what happens with OptReg and Redirector. While forcing fluxes to be above defined values forces precursors to be available regardless, fluxes regulated with TDPS can only carry flux if precursors are available (naturally or as a consequence of other regulations).

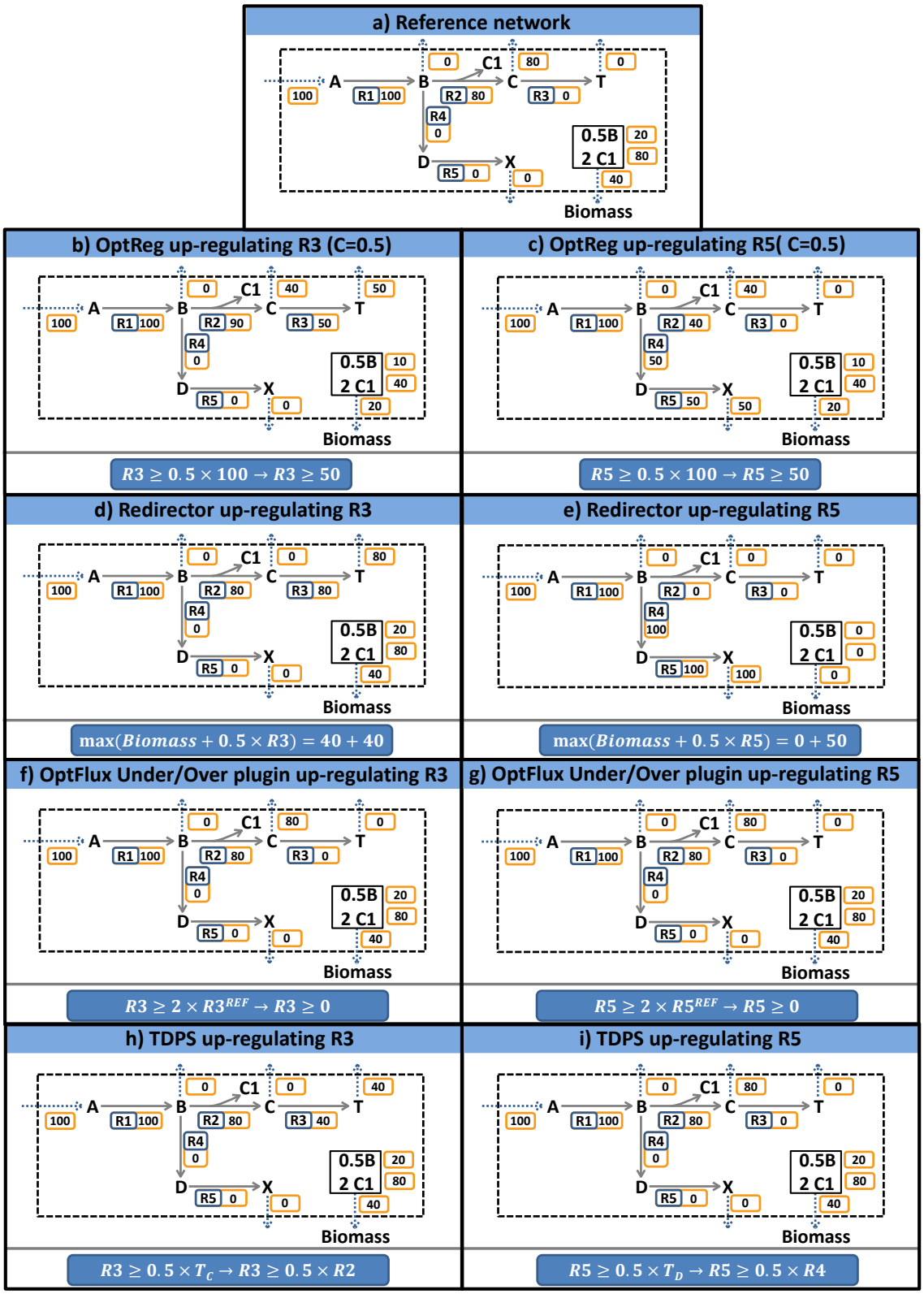

Figure S1 – Comparison of several methods to perform reaction up-regulation using genome-scale models. a) Flux distribution for the toy network obtained by maximizing biomass production; b) Flux distribution for the toy network obtained by up-regulating reaction R3 according to OptReg constraints (C=0.5); c) Flux distribution for the toy network obtained by up-regulating reaction R5 according to OptReg constraints (C=0.5); d) Flux distribution for the toy network obtained by up-regulating reaction R3 according to Redirector objective function constraints (objective weight for up-regulation assumed to be 0.5 in comparison to biomass weight); e) Flux distribution for the toy network obtained by up-regulating reaction R5 according to Redirector objective function constraints (objective weight for up-regulation

assumed to be 0.5 in comparison to biomass weight); f) Flux distribution for the toy network obtained by up-regulating reaction R3 with the under/over-expression plugin from OptFlux (assuming two times increase in comparison to the reference flux); g) Flux distribution for the toy network obtained by up-regulating reaction R5 with the under/over-expression plugin from OptFlux (assuming two times increase in comparison to the reference flux); h) Flux distribution for the toy network obtained by up-regulating reaction R3 with TDPS (assuming 50% of the reactant's turnover is consumed by R3); i) Flux distribution for the toy network obtained by up-regulating reaction R5 with TDPS (assuming 50% of the reactant's turnover is consumed by R5).

Supplementary Text S2- Algorithm flowchart for computing the flux partition ratios from a set of genetic modifications

TDPS is capable of simulating complex strain designs using two separate components: a newly developed modelling heuristic that can handle all types of genetic modifications (Figure 1 in main text) and an innovative objective function that promotes flux ratio rigidity after network disturbances (Figure 2 in main text). Figure S2 shows how these two components were integrated into a robust simulation algorithm designed specifically to allow quantitative phenotypical analysis of complex metabolically engineered strains. The first step in the simulation process is to choose a Genome Scale Metabolic Model (GEM) and define the environmental conditions by specifying a set of exchange fluxes (carbon source, nitrogen source, oxygen availability, etc.). These parameters are then used to compute a reference (wild-type) flux distribution using pFBA [4] as the simulation method. It is worthy of note that the reference flux distribution obtained with pFBA might change depending on the linear programming solver used. Since pFBA minimizes the total sum of fluxes in the network, assuming maximum biomass growth, it cannot determine a unique flux distribution if there are parallel pathways with equivalent stoichiometry in the network. Therefore, depending on the solver used, the reference flux distributions can have small variations in the fluxes through parallel pathways. In order to assure the reproducibility of the results obtained with TDPS, the reference flux distribution used should always be kept constant (the same is valid when other simulation tools that require a reference flux distribution are used, such as MOMA).

Using a reference flux distribution, the TDPS algorithm then starts the calculations necessary to obtain the normalized flux partition ratios ($X_{m,n}$) required to formulate the genetic modification

constraints and the objective function (Figure S2). To avoid confusing nomenclatures of the variables, the flux partition values used in the intermediary calculations shown in Figure S2 were named *F* values. First, the wild-type or reference flux partition ratios ($F^R_{m,n}$) are calculated using the reference flux distribution, by computing the fraction of the reference production turnover for metabolite *m* (T^R_m) that each of the consumer reactions (R_m) is using (Figure S2). The set of desired genetic modifications (*GM*) is then used to modify the $F^R_{m,n}$ values according to the regulation parameter *C* associated with each modification. If a reaction is active in the reference flux distribution ($F^R_{m,n} > 0$) and it is targeted for up-regulation ($1 < C \leq 5$), then the algorithm increases the fraction value for all its precursors (M_n) according to the formula shown in Figure S2. The up-regulation formula is composed of two terms: the first term increases the modified fraction value ($F_{m,n}$) up to four times the reference value and the second term further increases it by a 20% of the *C* parameter. If an active reaction is a target for down-regulation ($0 < C < 1$) then its $F_{m,n}$ value is decreased by direct multiplication with the *C* parameter. In both cases the magnitude of the *C* value should mimic the severity of the regulation applied to the organism and it is up to the user to find the value that best describes the genetic modification undertaken. Inactivating a reaction results in the $F_{m,n}$ values being set to zero.

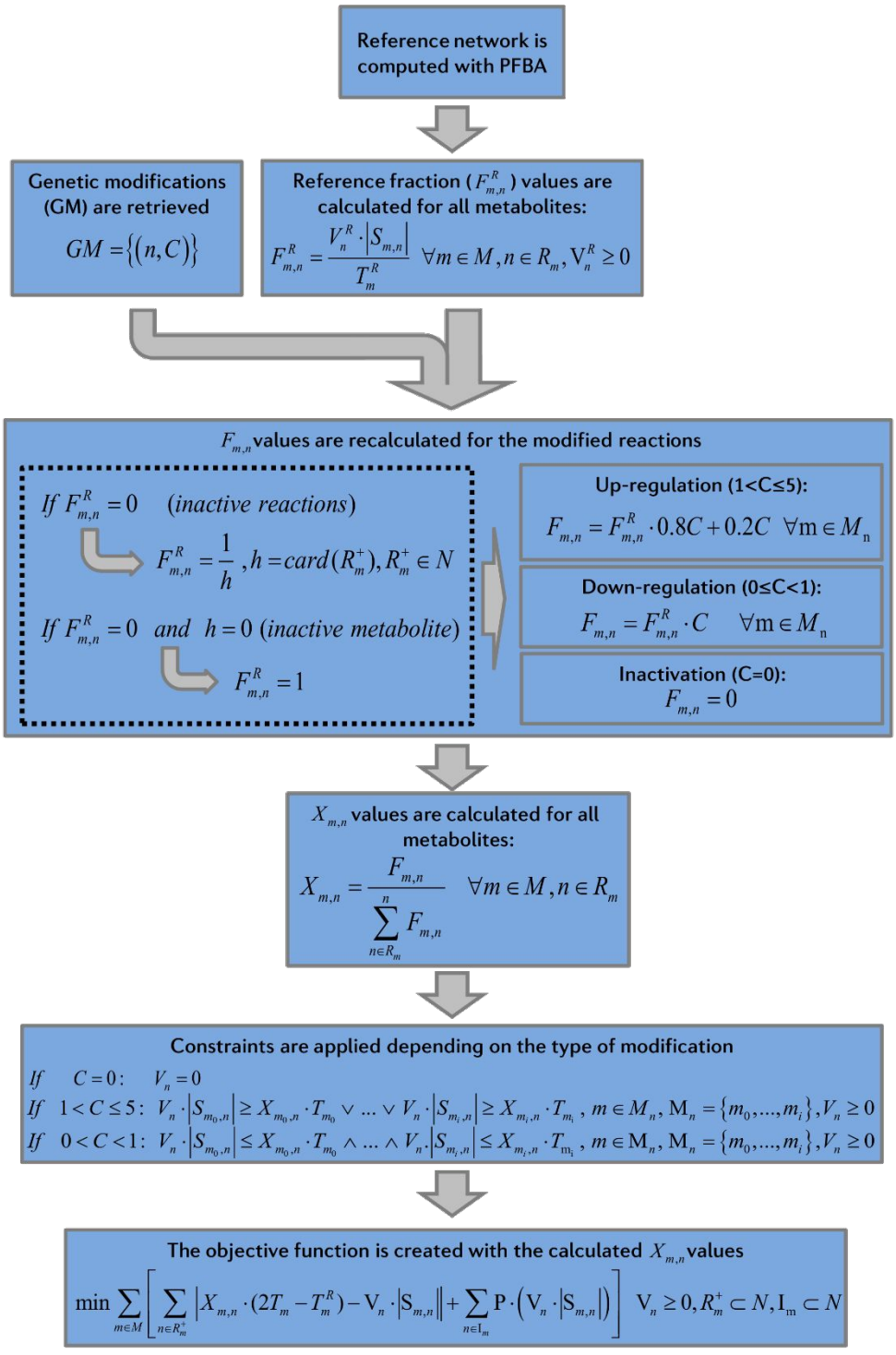

Figure S2- Flowchart describing the calculations performed by the TDPS algorithm (the detailed implementation is provided in the Material and Methods section). ***C***- regulation parameter, $F_{m,n}$- fraction of the production turnover consumed by reaction *n* (before normalization), $F^R_{m,n}$- fraction of the production turnover consumed by reaction *n* in the reference network, ***GM***- set of genetic modifications, ***h***- number of active reactions consuming metabolite *m* , I_m- set of inactive reactions in the reference flux distribution that can consume metabolite *m*, ***M***- set of all metabolites, M_n- subset containing all the precursors of reaction *n*, ***N***- set of all reactions, ***P***- penalty constant for activated reactions, R_m- set containing all the reactions that can consume metabolite *m*, R^+_m- set containing all the active reactions consuming metabolite *m* in the reference flux distribution, $S_{m,n}$-stoichiometric coefficient of metabolite *m* in reaction *n*, T_m- production turnover of metabolite *m* in the simulation, T^R_m- production turnover of metabolite *m* in the reference flux distribution, V_n- flux value of reaction *n* in the simulation, V^R_n- flux value for reaction *n* in the reference flux distribution, $X_{m,n}$- normalized fraction value.

When the reaction targeted for modification is inactive in the reference flux distribution ($F^R_{m,n} = 0$), it is necessary to estimate the fraction of the total production turnover that this reaction will consume when activated. TDPS uses the number of reactions that compete for the same substrate (h) to estimate the average $F^R_{m,n}$ value for each precursor required by the activated reaction. The number of reactions that compete for the same substrate (h) refers to the active consumers of metabolite m (R^+_m) in the reference flux distribution and is calculated using the cardinality (*card*) of R^+_m. With the estimated $F^R_{m,n}$ value, TDPS can then apply the same rules described above for up- and down-regulation, depending on the value of the C parameter. Although the up-regulation of an inactive reaction is quite straightforward to understand, it is also biologically relevant to down-regulate inactive reactions. For example, in heavily engineered strains it is very likely that some of the modifications applied will result in a regulatory response that induces the activation of reactions that are usually off. These activated reactions can be important down-regulation targets and it would not be appropriate to exclude them. Therefore, the estimated $F^R_{m,n}$ value is valid for both up and down-regulations.

There is also the possibility that an inactive reaction ($F^R_{m,n} = 0$) contains one or several precursors that are not being actively consumed by any reaction in the network ($h = 0$). In this case the TDPS algorithm assumes that the activated reaction will be the only consumer for those metabolites ($F^R_{m,n} = 1$).

After manipulating the $F_{m,n}$ values in accordance to the genetic modifications, the sum of all the flux partition ratios for a certain metabolite is no longer guaranteed to be equal to one. To normalize them back to unitary fractions, each $F_{m,n}$ is divided by the summed partition ratios for the respective metabolite, which yields the $X_{m,n}$ values (Figure S2). Using the rules described for the formulation of the turnover-dependent constraints, the $X_{m,n}$ values are then employed in the creation of flux constraints for all the reactions present in the *GM* set. Finally, the $X_{m,n}$ values are

also used in the formulation of the split ratio stability term of the objective function (Figure S2), which is then minimized by the solver in order to produce the mutant flux distribution.

Supplementary Text S3- Sensitivity analysis for the network rigidity objective function

The objective function of TDPS assumes that when a cell is disturbed the flux split ratios will stay as close as possible to a reference undisturbed state. Therefore, if the set of reference split ratios includes errors, the simulations performed with TDPS will be error prone. In order to verify how sensitive the production yields predicted with TDPS are to the rigidity oriented objective function, TDPS_FBA was used to test how much of the production yield variability is a consequence of the objective function used.

Two strains from the 3-HP case-study were used as benchmark for the sensitivity test: HPY01, a strain including the insertion of a single heterologous gene (malonyl-CoA reductase) and a significantly engineered strain (HPY11) that includes five modifications (three insertions and two up-regulations). These two strains were chosen because in HPY01 the disturbance cause by a single modification is small and in HPY11 the disturbances are more significative and the objective function has a much bigger role in the simulation. Figure S3a shows that for the strain HPY01, increasing the tolerance value has no effect whatsoever. Since the biomass is so close the optimum value, replacing the TDPS objective function with biomass maximization has no effect on 3-HP production. In contrast to these results, the sensitivity analysis for the strain HPY11 shows a much bigger role in both the biomass and 3-HP yields. By relaxing the network rigidity objective and maximizing biomass, the biomass yields progressively increase and approach the maximum value of 0.12 gCDW/gGlucose. With an opposite trend, the 3-HP production yield decreases with lower constraints on network rigidity.

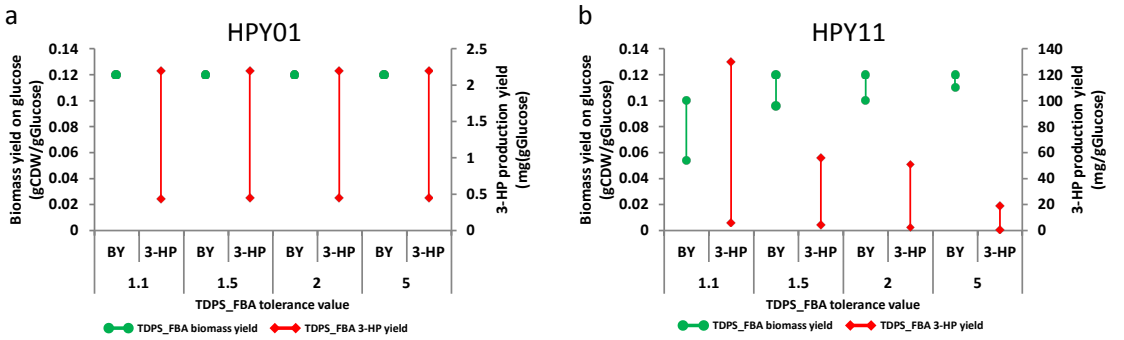

Figure S3- Sensitivity analysis of the TDPS simulations when the objective function is progressively relaxed. Simulations were performed with TDPS_FBA for the strain HPY01 (a) and HPY11 (b), using a tolerance constant of 1.1, 1.5, 2.0 and 5.0.

This analysis showed that the product yields simulated with TDPS are affected by the objective function used and as a consequence by the reference flux ratios used. Although this effect is less pronounced in simulations with small disturbances (strain HPY01), when more genetic changes are applied the simulated yields will be more dependent on rigidity objective function and the reference flux ratios used.

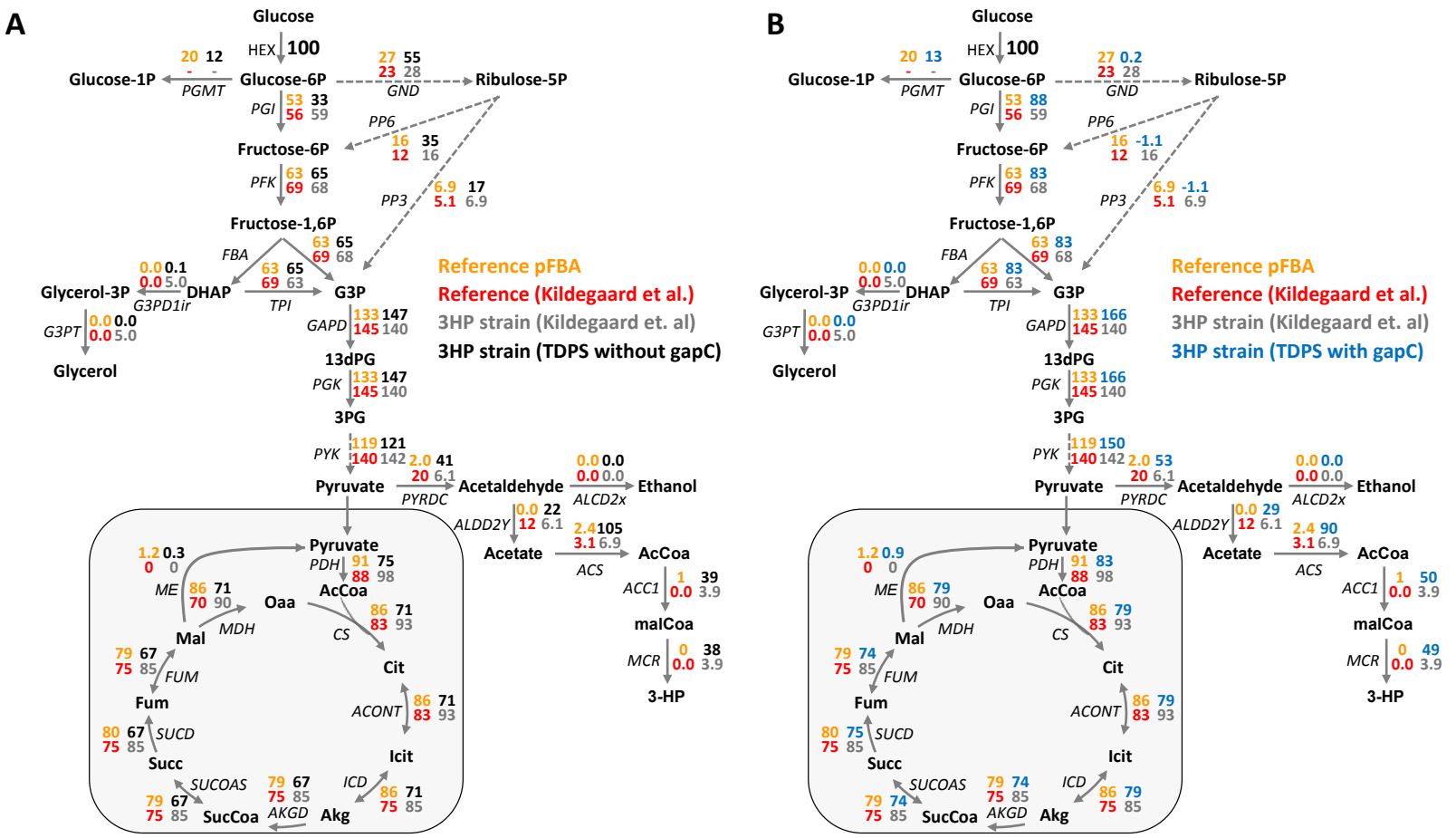

Figure S4 – Comparisson between the flux distributions simulated with TDPS for a 3-HP producing mutant and the fluxes determined by Kildergaard et. al using ^{13}C metabolic flux analysis. The measured flux distributions for a reference strain and a 3-HP producing mutant (down-regulations: Glyceraldehyde-3-phosphate dehydrogenase; insertions: *gapC* from *Clostridium acetobutylicum*, acetyl-CoA synthase derived from *Salmonella enterica* and *mcr* from *Chloroflexus aurantiacus*; up-regulations: *PDC1*, *ALD6* and *ACC1*) were obtained by Kildergaard et. al using ^{13}C metabolic flux analysis. The TDPS reference was obtained with pFBA and the TDPS flux distributions for the 3-HP producing mutant with (A) and without gapC insertion (B) were obtained assuming a C parameter of 2.0 for up-regulations and 0.50 for down-regulations. References: Kildegaard, K. R. *et al.* Engineering and systems-level analysis of *Saccharomyces cerevisiae* for production of 3-hydroxypropionic acid via malonyl-CoA reductase-dependent pathway. *Microb. Cell Fact.* **15,** 53 (2016).

References

(1) Pharkya, P., and Maranas, C. D. (2006) An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab. Eng. 8*, 1–13.

(2) Gonçalves, E., Pereira, R., Rocha, I., and Rocha, M. (2012) Optimization approaches for the in silico discovery of optimal targets for gene over/underexpression. *J. Comput. Biol. 19*, 102–14.

(3) Rockwell, G., Guido, N. J., and Church, G. M. (2013) Redirector: designing cell factories by reconstructing the metabolic objective. *PLoS Comput. Biol.* (Papin, J. A., Ed.) *9*, e1002882.

(4) Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., Adkins, J. N., Schramm, G., Purvine, S. O., Lopez-Ferrer, D., Weitz, K. K., Eils, R., König, R., Smith, R. D., and Palsson, B. Ø. (2010) Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol. 6*, 390.