* **Theoretical analysis about the kinetics equation form**

When I did Bayes inference simulation on the original version of the log-linear kinetics (Eqn. 1), it always got diverged and no regression of the kinetics parameters was got.



Fold change of enzymatic specific flux

Fold change of participated metabolites concentration

Elasticity coefficients of individual metabolites

(1)

As it always failed to do the Bayes inference on the elasticity coefficients, I turned back to the derivation of the kinetics and definition of elasticity coefficient. I found that there was mistake using of the elasticity coefficient in Eqn. 1. The elasticity coefficient is a local coefficient which is only valid around the reference steady state, i.e. the 0 condition in Eqn. 1. I tried to regress these parameters across a wide range of conditions (with specific growth rate ranging from 0.05 h-1 to 0.4 h-1), there is no way to get the coefficients with valid physiological meaning, even though I got regressed coefficients for some reactions by chance. Finally, I figured out that the equation is only valid under conditions (even dynamic conditions) near the reference condition, so the following equation was only valid as *vi* is not far from the steady state of S:

(2)

According to the linear relation between reaction rate and reaction affinity proposed by Onsager (Phys. Rev., 1931, 37:405-426), which is shown as follows:

(3)

Where L is phenomenological coefficient, and A is the reaction affinity (which equals to minus the change in free energy of the reaction). Here, we added the enzyme amount term (e) to the original equation, same expression form was also discussed in Visser 2003 (Metab. Eng., 2003, 5:164-176). It may be argued that the relation of eqn. 3 is only valid close to equilibrium, however, many empirical analysis observed that the linear relationship between reaction rate and reaction affinity is valid even if the reaction operates far from equilibrium (Rottenberg, Biophys. J., 1973, 13: 503–511; van der Meer, et al, Biochim. Biophys. Acta., 1980, 591: 488–493; Nielsen J, Biochem. J., 1997, 321:133-138).

The reaction affinity term was then substituted by the 2nd thermodynamic law equation as follows,

(4)

Where *Keq* is the reaction equilibrium constant and Q is the reaction quotient, take the following reaction as an example,

According to Eqn. 3 and 4, the net forward reaction rate can be expressed as follows,

(5)

Under one steady state c, the reaction rate *v* can be expressed as form of flux J. Divide the steady state flux by enzyme concentration will give an enzymatic specific flux j. under the specific steady state condition c, it gives,

(6)

It should be pointed out that *ai* is coefficient independent of the steady state conditions, and corresponding to allosteric effect of each metabolite. Furthermore, *ai* has the same dimension as *kcat*, the enzyme turn over number, thus we call these coefficients the **intrinsic turn over number** (*kcat,intrinsic\_i,* which is the intrinsic turn over number of enzyme for metabolite i) as respect to individual metabolite that take part in the reaction. With this concept we can even include allosteric effectors into Eqn. 6.

However, to apply Eqn. 6 in integrating multi-omics data, we also need to eliminate the equilibrium terms (*X\**) in the equation. We then take a specific steady state as reference state denoted by superscript 0, and the kinetics equation under reference state is as follows,

(7)

Subtract eqn. 7 from eqn. 6 we can get the final model we use to integrate fluxome, proteome and metabolome data. It can be expressed as follows,

(8)

The above thermokinetic equation separates the effect of enzyme and metabolite on the reaction flux by using enzymatic specific flux instead of reaction flux itself. It can be derived that difference of enzymatic specific flux under two conditions is determined by relative change of metabolite concentration and their corresponding kinetics parameters (**intrinsic turn over number**, *ai*).

* **MCMC Bayes inference of the intrinsic turn over number for each metabolite**

Given a reference state (with enzyme abundance (*e0*), fluxes (*J0*)), the linear thermokinetic equation （Eqn. 8） translates absolute enzyme abundance (*ec*), relative metabolite abundances with respect to reference (*Xc*/*X0*) and intrinsic turn over numbers (*kcat,intrinsic\_i*, i.e. *ai* in equation 8) into predicted flux () under condition *c*. To determine whether an investigated reaction obeys the above thermokinetic equation (Eqn. 8), we must find a set of kinetics parameters that best fit the FBA determined flux (). As there has no reports on the values of the proposed intrinsic turn over numbers, we want both maximum a posterior probability (MAP) estimator of *ai* and a measure of parameter uncertainty. To do this, we applied a Bayes inference approach to estimate these kinetic parameter *ai*:

(9)

The posterior distribution of parameter (*ai*) was estimated using Markov Chain Monte-Carlo based Bayes Inference, and the open source python Bayes package PyMC3 (Salvatier, et al, 2016) was used. A prior distribution of the model parameter Pr(*ai*) was first proposed, and samples of *ai* were drawn from the prior distribution, then, evaluating the reaction flux *Jsim* using these drawn *ai* through equation 8, following that a log-likelihood of *ai* () to determine how well the predict flux agree with the flux data (*Jobs*). Finally, the posterior probability of these drawn *ai* were calculated using equation 9 and it was determined whether the drawn *ai* should be accepted or rejected. Iteratively, the above steps drawn parameter from prior, evaluating predict flux, calculating log-likelihood and finally determine keeping or rejecting of drawn *ai* sample were repeated until the upper bound of iteration steps. All accepted *ai* form the posterior distribution space, which will tell the most probable value for each *ai* and the credibility interval.

Without any experience of the prior distribution of *ai*, we assume normal distribution for these parameters , i.e., with the mean equal to mean of observed turn over number and variance to be square of standard error of turn over number. FBA analysis combined with FVA was carried out using Yeast-GEM v 7.6, both point estimation and variation of fluxes were obtained. Point estimation of the flux was used to calculate the squared error between the observed flux and predicted one with equation 9 using *ai* drawn from the prior distrubiton. Like what Hackett (Hackett, et al., 2016) did, we assume the deviation between *jobs* and *jpred* followed a Normal distribution with variance given by the squared error. With the help of FVA the experimental uncertainty was introduced through flux variability. The log-likelihood of *ai* was then modified to account for the flux variability for estimation of posterior distribution of *ai*, the modified log-likelihood function was shown as follows,

(10)

is the cumulative distribution function of Normal distribution with parameter and . denotes the predicted specific flux under condition c, and and are upper and lower value of flux estimated using FVA. with n denotes number of experimental conditions and *I* denotes the number of metabolites involved in the model.

The log-likelihood function combined with prior distribution of model parameters were then used to calculate the posterior probability of the drawn *ai*, using MCMC algorithm the decision of dropping or keeping the drawn *ai* was made. Large iterative steps (20000 samples with discarding 1000 draws between consecutive two samples, so 120000 steps in total) were needed to guarantee the convergence, and value calculated by Gelman-Rubin statistic method (Gelman and Rubin, 1992) was used as criteria to confirm convergence of the inferenced parameters.

* **Strategy for inferencing the data-dependent-best fit model**

According to the definition of intrinsic turn over number of *ai*, the bigger of *ai* the more the corresponding metabolite will influence the enzyme’s *kcat*. Here, we assume that enzymes take as less metabolite as possible to influence their *kcat* value. With this assumption, we proposed the following strategy to get the plausible best fit model depending on the data in hand, what is called data-dependent-best fit model.

**ALGORITHM 1**: Inference of the data-dependent-best fit kinetic model for a bio-reaction

**Input**: FBA results of flux vector with component at each experimental condition (), absolute enzyme concentration vector at each condition (), FVA results upper and lower flux vector with components under each condition (), relative abundance of each metabolites take part in the reaction as respect to a reference condition (), candidate metabolite list that may have effect on the *kcat* (or in another word, the intrinsic turn over number of this metabolite can not be omitted)

**Output**: The “best” fit model in which the most possible list of *ai* were given, and their posterior distribution.

**Initialization**: Set all metabolites that take part in the reaction as candidate metabolite list (**Met**candidate); set current best fit model chosen metabolites (**Met**curr,best) list to be empty; set current best fit model **Mod**b,curr to be empty.

**LOOPing**: for each metabolite **m** in **Met**candidate do

set *ab* as already confirmed model paramters, for first round loop it is empty

set model structure (**Mod**m) as

do MCMC Bayes inference for model **Mod**m

do next step

Test using Leave One Out (LOO) cross-validation criteria to select best from among all **Mod**m, comparing the select out best model **Mod**m,best to **Mod**b,curr.

If Modm,best is not better than Modb,curr:

Return **Mod**b,curr

Else:

Append the best model corresponding metabolite m to **Met**curr,best

Remove m from **Met**candidate

Goto **LOOPing** do next round loop.

* **Results analysis**

Using the above MCMC Bayes inference algorithm, we do best-fitted model inference for each reaction step in the central carbon metabolism first (Table 1). There are 28 reaction steps in the central carbon metabolism pathway (10 from EMP, 10 from TCA and 8 from HMP), and it was found that among all the 9 studied dilution rates, FBA results showed flux through the reaction step catalyzed by transaldolase in HMP were all zero, even though the corresponding enzyme protein were detected under all conditions. When doing Bayes inference modelling, this reaction step was not considered as no reaction flux information for it. The above Bayes inferences strategy (ALGORITHM 1) were carried out on the other 27 reactions, and 20 of them got fitted using the proposed model (equation 9), and the best fitted model structures were inferenced based on the observed fluxome, proteome and metabolome data, respectively.

**Dealing with missing data of metabolome**: Among all the test 28 reactions, not all metabolites of each reaction were detected in our metabolome data. As it cannot distinguish water generated in reaction and that in the solution, relative change of water amount was not considered in our model. As all experiment were carried out under steady state, CO2 level in system was considered to be constant and quick equilibrium between extra- and intra-cellular level of CO2 was believed to be the case, thus CO2 was also not considered in our model. We also assume stable intracellular pH level across all conditions, so there is no change of intracellular H+ level. According to Suarez-Mendez, 2016, Met. Eng. Comm., 3, 52-63, a linear relationship between the intracellular ATP level and specific growth rate is assumed. Reaction GPI (glucose-6-phosphate isomerase) and TPI (triose-phosphate isomerase) are considered to be under equilibrium, and *Keq* values of 0.259 (GPI) and 0.039 (TPI) taken from data in Canelas, et al, 2011, Met. Eng., 13:294-306 are applied to get abundance information of F6P and GAP. Other missing data of metabolites are taken as no-change across all conditions. Missing metabolites for each tested reaction step were listed in Table 1.

Table 1. Reactions (central carbon metabolism) considered in the Bayes inference modelling. Missing data on metabolites and reported regulators are listed for each reaction.

|  |  |  |  |
| --- | --- | --- | --- |
| **Pathway** | **Reaction** | **Missing data on metabolites[[1]](#footnote-1)** | **Regulators reported[[2]](#footnote-2)** |
| EMP | Hexokinase | ATP |  |
| glucose-6-phosphate isomerase | F6P |  |
| phosphofructokinase | ATP | 3PG(-),isoCIT(-), PEP(-),FBP(+) |
| fructose-bisphosphate aldolase | GAP, | ADP(-), AMP(-), CIT(-), E4P(-), dATP(+), HISnol(+) |
| triose-phosphate isomerase | GAP | AMP(-), GTT(-), GTTdS(-), ASP(+) |
| glyceraldehyde-3-phosphate dehydrogenase | GAP |  |
| phosphoglycerate kinase | 13PG, ATP |  |
| phosphoglycerate mutase | - |  |
| enolase | - |  |
| pyruvate kinase | ATP | MAL(-), aKG(-), 3PG(-), AMP(-), CIT(-), isoCIT(-), PEP(-), FBP(+), 6PG(+), DHAP(+), GAP(+), ALA(+), ASP(+), GLU(+), GLY(+), ILE(+), MET(+), R5P(+) |
| HMP | glucose 6-phosphate dehydrogenase | NADP, 6PGL, NADPH | AMP(-), PEP(-) |
| 6-phosphogluconolactonase | 6PGL |  |
| phosphogluconate dehydrogenase | NADP, Ru5P, NADPH | ASP(+), HIS(+), TRP(+) |
| ribulose 5-phosphate 3-epimerase | Ru5P, Xu5P, |  |
| ribose-5-phosphate isomerase | Ru5P |  |
| transketolase 1 | Xu5P |  |
| transaldolase[[3]](#footnote-3) | E4P |  |
| transketolase 2 | E4P, Xu5P |  |
| TCA | pyruvate carboxylase | ATP, ADP, H+, OA |  |
| citrate synthase | H+ |  |
| citrate to cis-aconitate(3-) | Cis-CAN |  |
| cis-aconitate(3-) to isocitrate | Cis\_CAN |  |
| isocitrate dehydrogenase | - |  |
| oxoglutarate dehydrogenase | H+ |  |
| succinate-CoA ligase | SUC-CoA |  |
| succinate dehydrogenase | SUC-CoA, ubQ |  |
| fumarase | - |  |
| malate dehydrogenase | H+ | CIT(-) |

According to the best fitted model searching algorithm (ALGORITHM1), there will be one best fitted model that include minimum amount of plausible allosteric effectors (here we only consider the metabolites that take part in the reaction, no other allosteric regulators were considered), e.g. For the first reaction step of EMP (the HX catalyzed reaction), the best fitted model shows that it is plausible ADP and ATP have allosteric effect on the enzyme (the HX enzyme), while allosteric effect of glucose and G6P can be omitted. The inferenced intrinsic turn over numbers of best fitted model for each of the 20 reactions are plotted in Figure 1. It can be seen that not all participated metabolite show allosteric regulation effect for each reactions, however, ADP、ATP and phosphate show regulating the *kcat* value for all reactions that they take part in, 7 among all the tested 27 reactions. The same conclusion is also observed for NADH for all 4 reactions it takes part in. Positive *ai* value means positive regulation of the corresponding metabolites to the enzyme *kcat* value, and vice versa. It is more confident to conclude that ADP shows positive effect on reaction HK and reaction PYC, NADH on reaction GAPD, ATP on reaction PGK, CIT on reaction CITS, while negative effect of PEP on reaction ENO, and not so strong confident on negative effect of DHAP on reaction FBA and TPI.

High accurate Bayes inference results depends on both the proposed model structure (whether we have chosen the right model for describing the observed data) and the amount of information about the system we are studying (whether enough of data for accurately inferencing the proposed model). Here, the wide 95% credible interval of the results in Figure 1 may be caused by two reasons: first, for some reactions we did not get all participated metabolite’s abundance information, which will make the inferenced *ai* value contains the contribution of non-measured metabolites (that makes the uncertainty higher for the inferenced *ai* value). Second, here we only considered 9 chmostat conditions, which only leave us 8 available data for inferencing (one condition is chosen as reference), lack of observed data also limit our ability to get accurate inference for the intrinsic turn over numbers for each reaction. However, using the LOO criteria, the most possible kinetics model structure is figured out by the limited data, so we call the results as data-dependent-best fitted model inference. And the regulation effect of ATP, ADP and NADH has been highly recommended by our model inference results, which makes sense.

Using the data-dependent-best fitted kinetics models that are inferenced using the Bayes modeling, we can simulate the corresponding enzymatic specific fluxes under all tested conditions for each reactions (we call it enzymatic specific flux simulated by Bayes inferenced model). Comparison between the simulated flux and that observed by FBA is shown in Figure 2. Even though the now fitted models do not show accurate prediction to the enzymatic specific fluxes, some of the models show correct changing trend of fluxes. Here we didn’t use the Pearson correlation coefficient to calculate the R2 like that in (Hackett, 2016, Science), instead, we use the Bayes R2 score value which take consideration of variance of posterior distribution of the predicted fluxes and showed better description of the goodness of fit. Another finding about the Bayes inferenced model is that, the model underestimate most reactions. This may imply that there are other effectors we did not considered in the model. Hackett, et al, showed that PFK had 4 regulators, FBA had 6, GAPD had 4, PGK had 20, GPD had 2, PGD had 3 and FUM had 1, details were listed in Table 1.

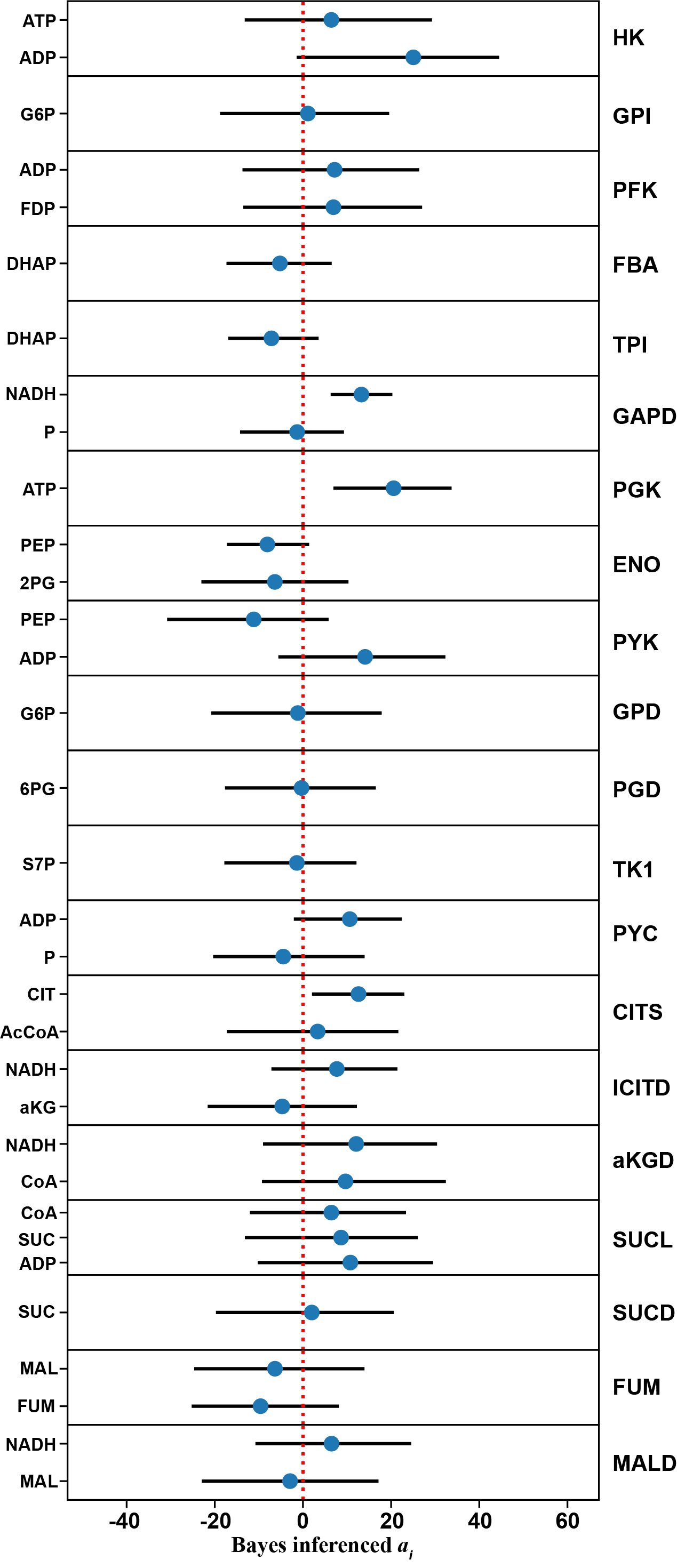
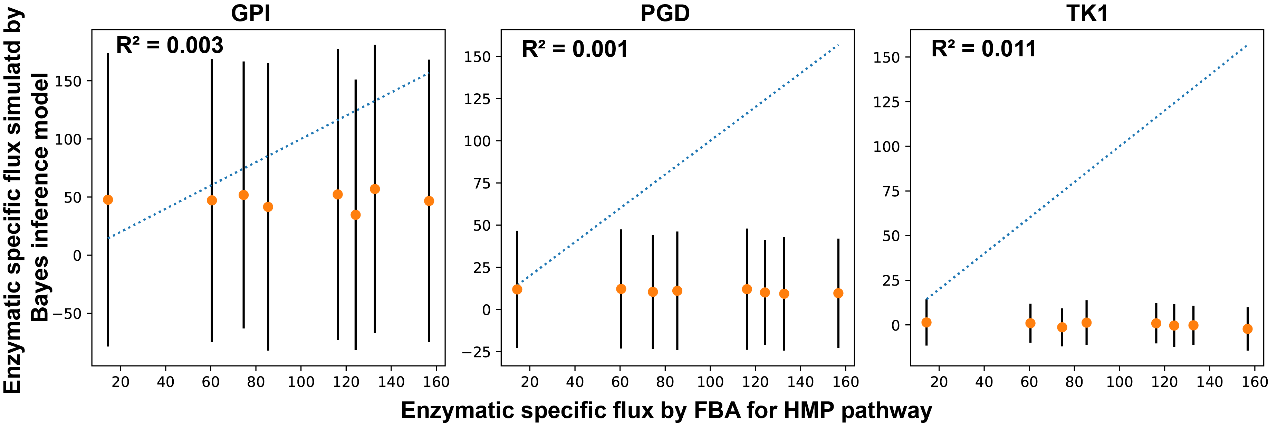
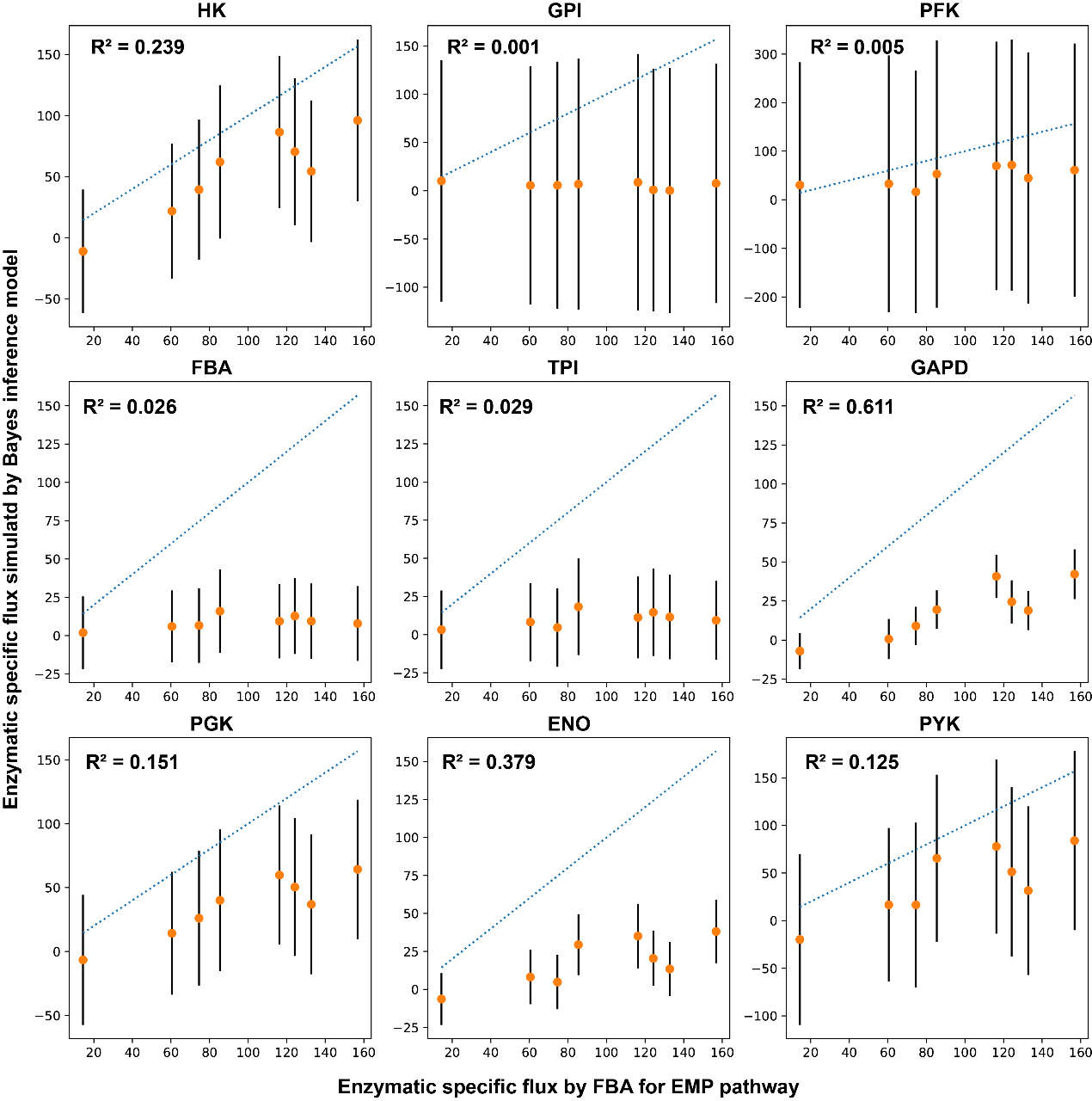


Figure 1. Bayes inference results of best fitted model for central carbon metabolism pathways. Filled circle denotes mean of the posterior distribution of each intrinsic turn over number *ai*from the best fitted kinetics model, and bars across each circle represents the 95% credible-interval gotten from the posterior distribution of *ai*. The red vertical dotted line located at 0 is shown as a reference, *ai* value higher than 0 means positive effect on the *kcat* of enzyme, and vice versa. Abbreviations: HK: hexokinase, GPI: glucose-6-phosphate isomerase, PFK: phosphofructokinase, FBA: fructose-bisphosphate aldolase, TPI: triose-phosphate isomerase, GAPD: glyceraldehyde-3-phosphate dehydrogenase, PGK: phosphoglycerate kinase, ENO: enolase, PYK: pyruvate kinase, GPD: glucose 6-phosphate dehydrogenase, PGD: phosphogluconate dehydrogenase, TK1: transketolase 1, PYC: pyruvate carboxylase, CITS: citrate synthase, ICITD: isocitrate dehydrogenase (NAD+), aKGD: oxoglutarate dehydrogenase (dihydrolipoamide S-succinyltransferase), SUCL:succinate-CoA ligase (ADP-forming), SUCD: succinate dehydrogenase (ubiquinone-6), FUM: fumarase, MALD: malate dehydrogenase.



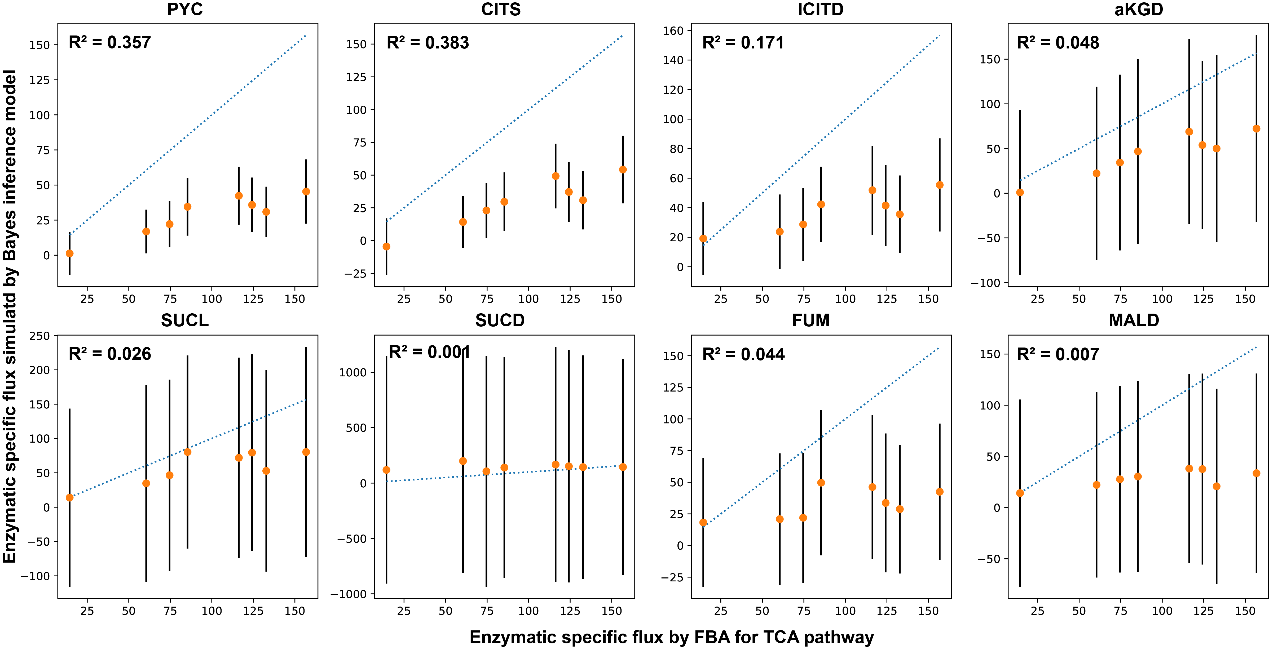


Figure 2. Comparison between simulated flux by Bayes inferenced model and FBA for the central carbon metabolism pathways. Abbreviations: the same as Figure 1. R2 value was calculated using the Bayes R square score definition, which equals variance of simulated flux (Varpred) over sum of variance of observed flux (Varobs) and Varpred, i.e. . By this definition, R2 value is restrict to then range of [0, 1], the higher the R2 value, the better the model fit.

**Reference:**

Gelman A., and Rubin D.R., (1992) A single series from the Gibbs sampler provides a false sense of security. In Bayesian Statistics (eds. J. Bernardo et al.), Oxford University Press, 625-631.

Hackett S.R., et al., (2016) Systems-level analysis of mechanisms regulating yeast metabolic flux, Science, 354, aaf2786.

Nielsen J, (1997) Metabolic control analysis of biochemical pathways based on a thermodynamic description of reaction rates, Biochem. J., 321:133-138

Onsager L., (1931) Reciprocal relations in irreversible processes. Phys. Rev. 37, 405–426.

Rottenberg H., (1973) The thermodynamic description of enzyme catalyzed reactions. The linear relation between the reaction rate and the affinity. Biophys. J. 13, 503–511.

Salvatier J., Wiecki T.V., Fonnesbeck C. (2016) Probabilistic programming in Python using PyMC3. PeerJ Computer Science 2:e55 DOI: 10.7717/peerj-cs.55.

van der Meer R., Westeroff H.V., van Dam K., (1980) Linear relation between rate and thermodynamic force in enzyme catalyzed reactions. Biochim. Biophys. Acta. 591, 488–493.

Visser D. and Heijnen J.J. (2003) Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. Metabolic Engineering 5(3): 164-176.

1. H2O and CO2 data was not considered. [↑](#footnote-ref-1)
2. Hackett, et al, 2016, Science, [↑](#footnote-ref-2)
3. FBA results of flux for this reaction are zero under all conditions. [↑](#footnote-ref-3)