

Supplement

Methods

Amino Acid Biosynthesis Pathways. If multiple amino acids are synthesized in the same pathway, steps common to the two or more amino acids have been left out. For example, if a precursor P leads to synthesis of two amino acids X and Y via a common intermediate A, we have not considered the regulation of A and only considered the regulation of reaction from intermediates C and D for synthesis of X and Y respectively (Fig. S10A). In addition, if an amino acid is synthesized by multiple pathways then unique steps involved in all pathways have been considered as shown in (Fig. S10B). (Sheet S4 in Excel).

Mathematical Analysis of the six Topologies. The equations describing the dynamics of each topology are as below.

Topology A

$$\frac{d[R]}{dt} = b - k_f \cdot [R] \cdot [S] + k_r \cdot [R^*] - k_{dR} \cdot [R]$$

$$\frac{d[R^*]}{dt} = k_f \cdot [R] \cdot [S] - k_r \cdot [R^*] - k_{dR} \cdot [R^*]$$

$$\frac{d[T]}{dt} = \frac{k_T \cdot [R^*]^2}{[\theta_T]^2 + [R^*]^2} - k_{dT} \cdot [T]$$

Topology B

$$\frac{d[R]}{dt} = b + \frac{k \cdot [R^*]^2}{[\theta]^2 + [R^*]^2} - k_f \cdot [R] \cdot [S] + k_r \cdot [R^*] - k_{dR} \cdot [R]$$

$$\frac{d[R^*]}{dt} = k_f \cdot [R] \cdot [S] - k_r \cdot [R^*] - k_{dR} \cdot [R^*]$$

$$\frac{d[T]}{dt} = \frac{k_T \cdot [R^*]^2}{[\theta_T]^2 + [R^*]^2} - k_{dT} \cdot [T]$$

Topology C

$$\frac{d[R]}{dt} = b + \frac{k}{1 + \left[\frac{R^*}{\theta} \right]^2} - k_f \cdot [R] \cdot [S] + k_r \cdot [R^*] - k_{dR} \cdot [R]$$

$$\frac{d[R^*]}{dt} = k_f \cdot [R] \cdot [S] - k_r \cdot [R^*] - k_{dR} \cdot [R^*]$$

$$\frac{d[T]}{dt} = \frac{k_T \cdot [R^*]^2}{[\theta_T]^2 + [R^*]^2} - k_{dT} \cdot [T]$$

Topology D

$$\frac{d[R]}{dt} = b - k_f \cdot [R] \cdot [S] + k_r \cdot [R^*] - k_{dR} \cdot [R]$$

$$\frac{d[R^*]}{dt} = k_f \cdot [R] \cdot [S] - k_r \cdot [R^*] - k_{dR} \cdot [R^*]$$

$$\frac{d[T]}{dt} = \frac{k_T}{1 + \left[\frac{R}{\theta_T} \right]^2} - k_{dT} \cdot [T]$$

Topology E

$$\frac{d[R]}{dt} = b + \frac{k \cdot [R]^2}{[\theta]^2 + [R]^2} - k_f \cdot [R] \cdot [S] + k_r \cdot [R^*] - k_{dR} \cdot [R]$$

$$\frac{d[R^*]}{dt} = k_f \cdot [R] \cdot [S] - k_r \cdot [R^*] - k_{dR} \cdot [R^*]$$

$$\frac{d[T]}{dt} = \frac{k_T}{1 + \left[\frac{R}{\theta_T} \right]^2} - k_{dT} \cdot [T]$$

Topology F

$$\frac{d[R]}{dt} = b + \frac{k}{1 + \left[\frac{R}{\theta}\right]^2} - k_f \cdot [R] \cdot [S] + k_r \cdot [R^*] - k_{dR} \cdot [R]$$

$$\frac{d[R^*]}{dt} = k_f \cdot [R] \cdot [S] - k_r \cdot [R^*] - k_{dR} \cdot [R^*]$$

$$\frac{d[T]}{dt} = \frac{k_T}{1 + \left[\frac{R}{\theta_T}\right]^2} - k_{dT} \cdot [T]$$

In the above equations, b (nM/s) is a rate constant of basal production of R. k_f ($s^{-1} \cdot nM^{-1}$) and k_r (s^{-1}) are association and dissociation constants of Regulator to signal respectively. k and k_T (nM/s) are maximal expression level of Regulator-dependent production of R and T respectively. θ and θ_T (nM) are concentrations of activator/repressor yielding half-maximal expression of R/R* and T respectively whereas k_{dR} and k_{dT} (s^{-1}) are degradation coefficients of R/R* and T respectively.

Generations of networks from a topology. In Topologies A, B, and C, expression of target, T is controlled by active regulator, R* in presence of signal whereas in topologies D, E, and F, inactive regulator, R controls the expression of target, T in absence of signal. We assumed the same target, T in all six topologies to study how the structure of control design confers different behavior to target T dynamics. Therefore, the degradation of T was assumed at same rate in all six topologies. The regulator, R gets converted into induced regulator, R* in the presence of signal. The conversion of R into R* was considered constant in all six topologies. For topologies A and D, four parameters (b , k_{dR} , k_T & θ_T) were chosen in the mathematical representation. Eleven uniformly spread parameter values were chosen for each parameter, and permutation for each resulted in the 11^4 (14641) networks. Similarly, five different values were chosen for the

six parameters (b , k_{dR} , k , Θ , k_T & Θ_T) describing kinetics in B, C, E, and E, resulting in 5^6 (15625) networks.

Results from simulation

Dynamics of activation and deactivation vary across the six topologies. Analysis of time of activation and deactivation indicate that subtle differences exist in the T-t50 space covered by six topologies (Figure S5). In the ON state, topology C and F exhibit the widest range of steady state T and activation time. This suggests that topologies C and F are more suited to speed up the activation of targets with small steady state levels and it is well known fact that the negative auto-regulation speeds up the cellular responses. In the OFF state, we note that physiological functions with preference for smaller steady state T values and higher deactivation times would have a greater chance to be represented by topologies A-C. We also applied this analysis on all the networks from six topologies irrespective of the “Performance Box”. In (Figure S6), we observe that the topology E alone demonstrates high activation time of targets with smaller steady state levels and topology F stands out to show widest range of steady state and activation time. We speculate that this is one of the reasons of over-representation of topology F, and under-representation of topology E in natural circuits.

Cost of control of expression. For all networks in “Performance Box”, change in steady state target response (Δ Target) in both switching ON (blue) and OFF (red) against change in amount of total regulators (Δ Regulator) at steady state have been plotted in (Figure S7). Total regulators include both active and inactive form of regulator. Positive value of Δ Target indicates the switching of a network from OFF to ON condition whereas negative value indicates switching from ON to OFF condition. Positive value of Δ Regulator indicates that a particular switching

needs production of more regulators and negative value of $\Delta\text{Regulator}$ indicates regulators need to be degraded for that particular switching. Networks in topologies A-C are designed to remain expressed at most of the time whereas topologies D-F are designed to remain under repression most of the time. Intuitively, networks in topologies A-C should offer qualitatively similar $\Delta\text{Target}-\Delta\text{Regulators}$ profile. Likewise, networks in topologies D-F should have similar $\Delta\text{Target}-\Delta\text{Regulators}$ profile qualitatively. However, in (Figure S7), we note that topologies A and D, topologies B and F, and topologies C and E offer qualitatively similar $\Delta\text{Target}-\Delta\text{Regulator}$ profile. In topologies A and D, we observe that ΔTarget values for all networks are plotted at almost zero change in total regulators. This is due to absence of feedback. Target expression in A and D topologies are controlled by only inter-conversion between active and inactive regulators, hence amount of total regulators is roughly constant. This suggests that these topologies have the capability to offer cheapest but poor control over expression of T. This likely confers a disadvantage to cells when they move from one environmental condition to another. The $\Delta\text{Target}-\Delta\text{Regulator}$ plane in (Figure S7) shows that topologies C and E offer dynamics where production of regulators is required to switch the network OFF and degradation of regulators is required to switch the network ON. On other hand, topologies B and F offer a dynamics where they require production of regulators to switch the network ON and degradation of regulators during switching network OFF (Figure S7). In other words, it suggests that topologies C and E are expensive in switching OFF whereas topologies B and F are expensive in switching ON. As we know that activator-based topologies (A, B, C) are designed to be expressed frequently, hence their control cost in switching ON should be minimized and repressor-based topologies (D, E, F) are designed to be expressed less frequently (remain repressed frequently) and thus should offer minimum cost and precise control over switching OFF. Topologies C and F are the only topologies which provide economical switching ON and OFF respectively and also demonstrate continuous spread of networks over $\Delta\text{Target}-\Delta\text{Regulator}$ plane (Figures S7). In (Figure S8), we show that how economic a control in a

particular topology is to maintain a certain response level at steady state. Our analysis (Figures S8) suggests that topologies A-C and topologies D-F are qualitatively similar in showing maintenance cost dynamic. However, topology B and C, for some networks, require high number of regulators to maintain target response in ON and OFF conditions respectively. Again, topology E only, in both ON and OFF conditions, stands expensive by requiring high number of regulators to maintain response level at steady state. In terms of control of switching and maintaining response level, topology F is suited for most of the cellular functions in the cell whereas topology E stands the most expensive.

Switchability and sensitivity. For optimal functioning, *E. coli* would need to control expression of a large number of genes. However, the control and tuning of each gene in the OFF or ON state will be unique. Hence, each physiological role would require a topology most suited to provide appropriate control. From our analysis (Figure S9, Left Panel), we note that the response dynamics of networks in topologies A-C are qualitatively similar whereas response dynamics of networks in topologies D-F are qualitatively similar. The switchability of networks in topologies A-C is more effective than networks in D-F but they offer very limited response dynamics. The networks in topologies D-F are suited to provide wide range of response dynamics. In addition, topology F offers the widest and continuous spread ratio of steady state T in ON and OFF conditions, across a large range of T values. Similar observation has been noted in sensitivity test for all topologies (Figure S9, Right Panel). The networks in topologies D-F are more sensitive to signal strength and they offer wide range of response dynamics with varying signal strength. Topology F offers the widest and continuous space over ratio of steady state T in high and low signal strength, and hence is likely most suited for most cellular functions.

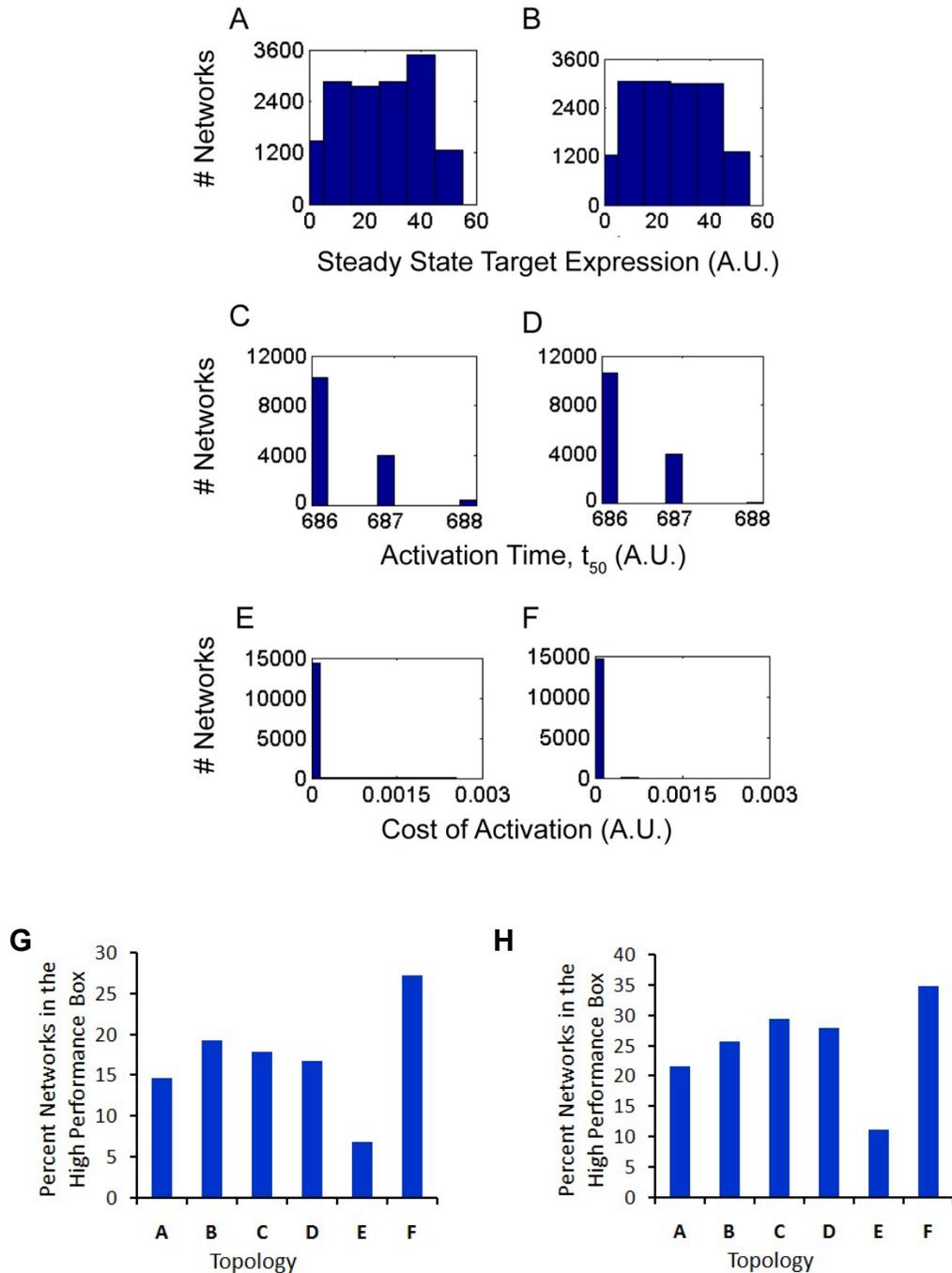


Figure S1. Comparison between our method and latin hypercube method of parameter sampling. The networks from topology A have been distributed based on their (A, B) steady state target expression, (C, D) activation time, and (E, F) cost of activation. (A, C, E) are from our method of parameter sampling and (B, D, F) are from latin hypercube method of parameter sampling. (G) Frequency distribution in the "High Performance Box" when target T is being expressed at a basal rate, b . (In the simulations b was taken as 1/100 of the maximal rate of

target production). (H) Frequency distributions of topologies in the "High Performance Box" when parameter values are spread in a log-normal manner over the defined range.

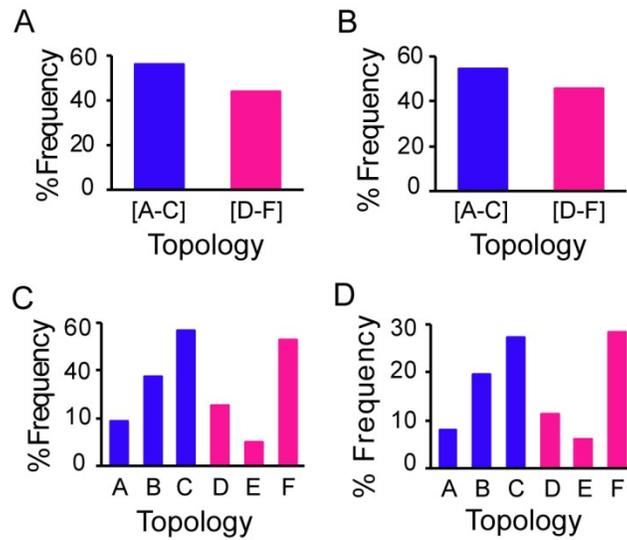


Figure S2. Frequency distribution of R-T interactions in *E. coli*. All (A and C) Regulator-Gene interactions (B and D) Regulator-Promoter interactions in *E. coli*, with including global regulators, were calculated from RegulonDB and then distributed (A and B) among activator-[A-C] and repressor-based [D-F] topologies and (C and D) among the six distinct topologies.

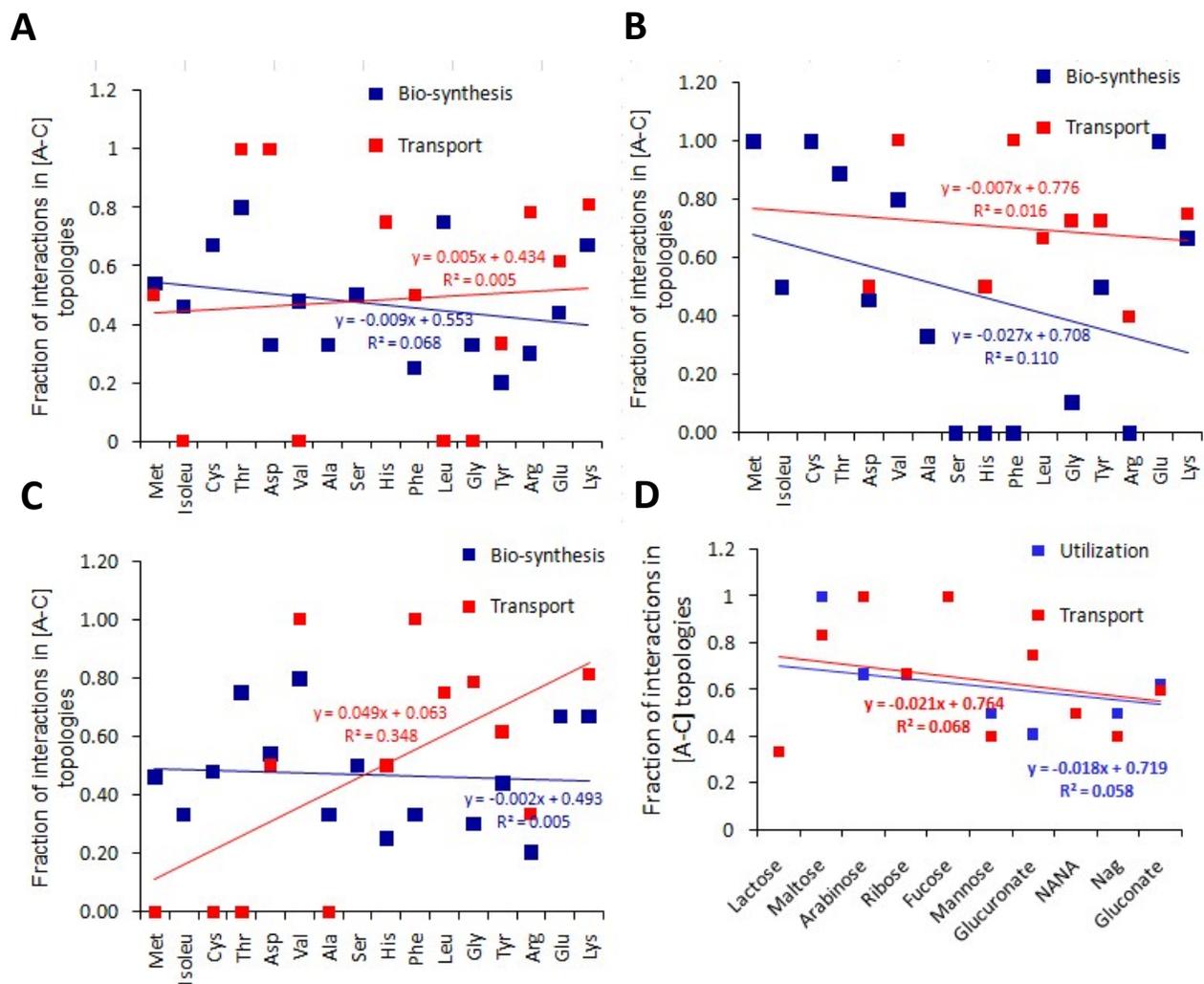


Figure S3. Correlation between demand and activator-based control. (A) X-axis represents amino acids in increasing availability to *E. coli*, and y-axis represents fraction of regulatory interactions in the activator-based topologies, while global regulators were included, for each amino acid biosynthesis regulon (blue) and its transport (red). (B) X-axis represents amino acids in increasing availability to *E. coli* (taking into account relative expression of proteins in the *E. coli* genome), and y-axis represents fraction of regulatory interactions in the activator-based topologies for each amino acid biosynthesis regulon (blue) and its transport (red). (C) X-axis represents amino acids in increasing availability to *E. coli* (taking into account relative expression of proteins in the *E. coli* genome), and y-axis represents fraction of regulatory interactions in the activator-based topologies, while global regulators were included, for each amino acid biosynthesis regulon (blue) and its transport (red). (D) X-axis represents carbon sources in increasing order of preference to *E. coli* in the intestine, and y-axis, the fraction of regulatory interactions in the activator-based topologies, in presence of global regulators, controlling expression of metabolic genes involved in utilization (blue) of each sugar.

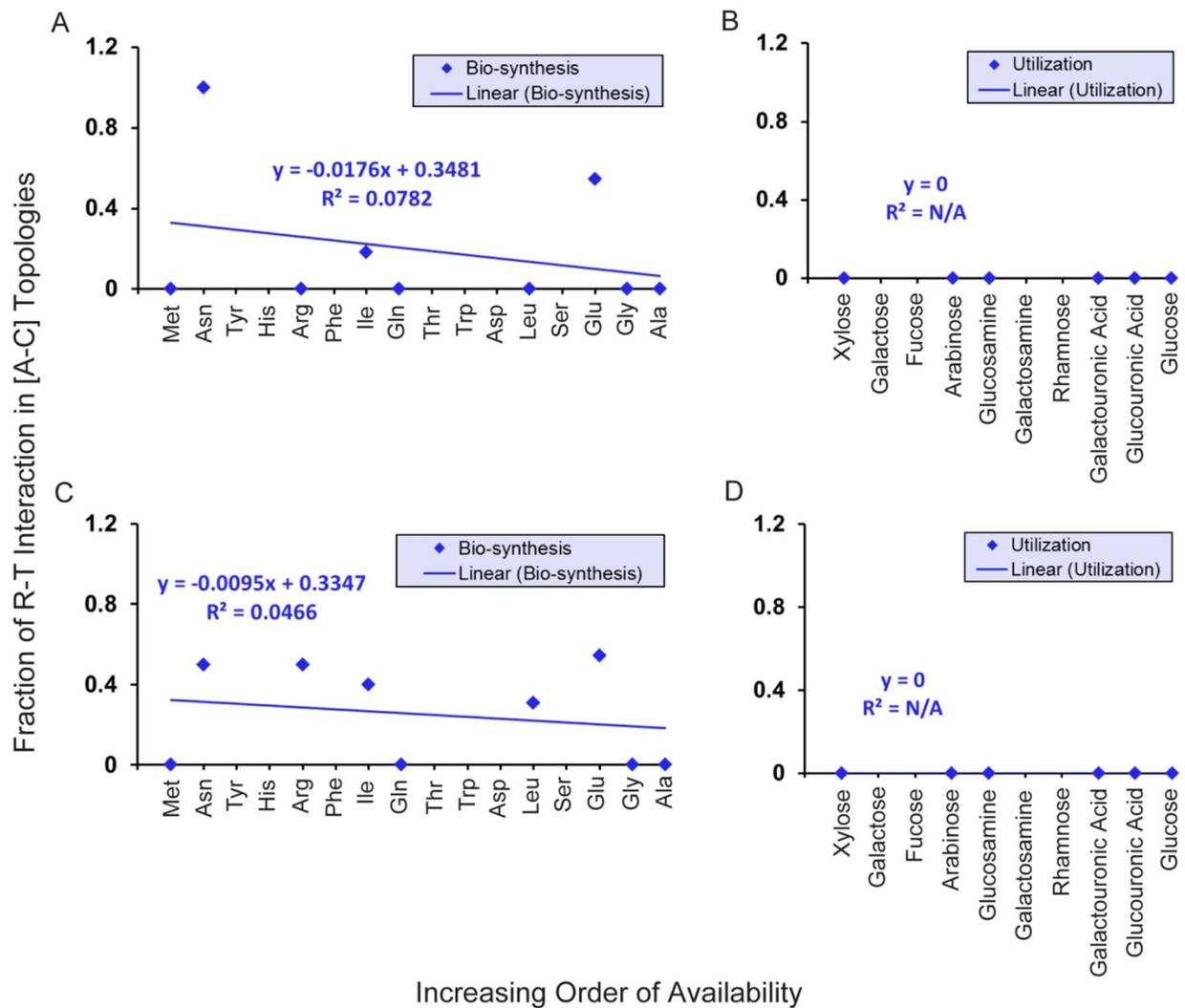


Figure S4. Correlation between demand and activator-based control.(A and C) X-axis represents amino acids in increasing availability to *B. subtilis*, and y-axis represents fraction of regulatory interactions in the activator-based topologies for each amino acid biosynthesis regulon (blue). (B and D) X-axis represents carbon sources in increasing availability to *B. subtilis* in the soil, and y-axis, the fraction of regulatory interactions in the activator-based topologies controlling expression of metabolic genes involved in utilization (blue) of each sugar. (A and B) Global regulators were excluded in the calculation of regulatory interactions. (C and D) Global regulators included in the calculation of regulatory interactions.

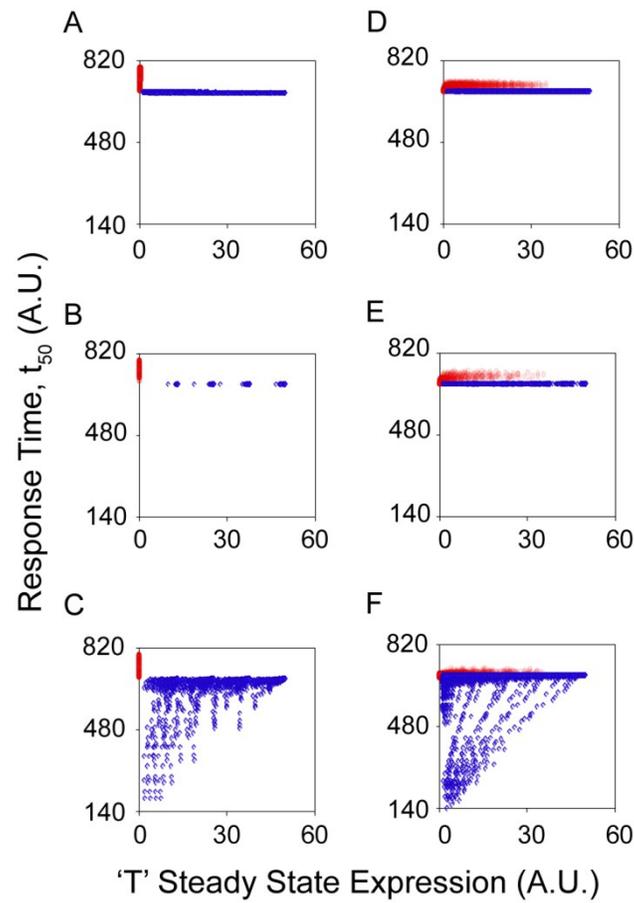


Figure S5. Response time in switching the topology ON and OFF. Response time is the time required to reach at 50% of the steady state expression of target. Response time of networks as blue dots in ON condition and red dots in OFF condition have been plotted with steady state T expression.

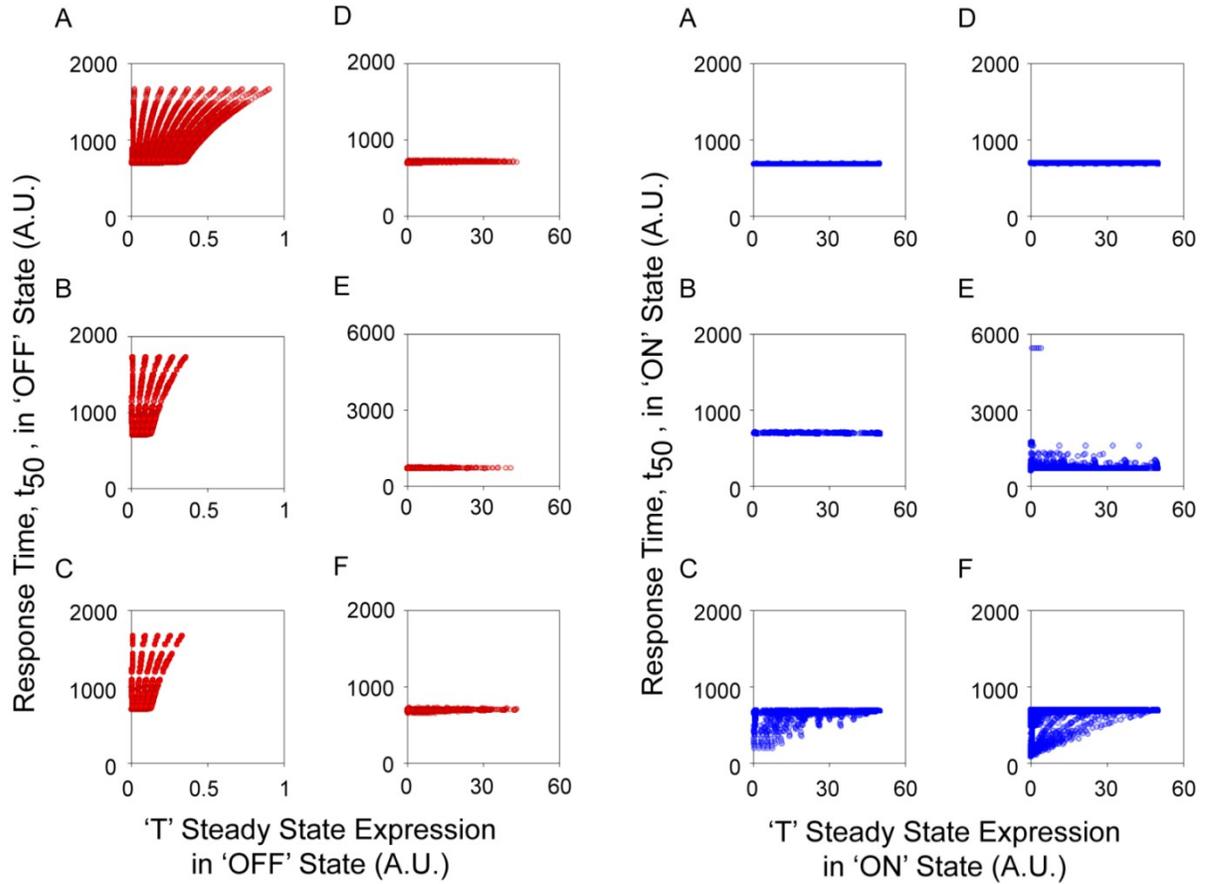


Figure S6. Response time in switching ON and OFF. All networks (~90000) were included in this plot. Blue and red dots indicate activation and deactivation time respectively against their steady state response. Topologies E and F out stand all to slow down and speed up the response respectively.

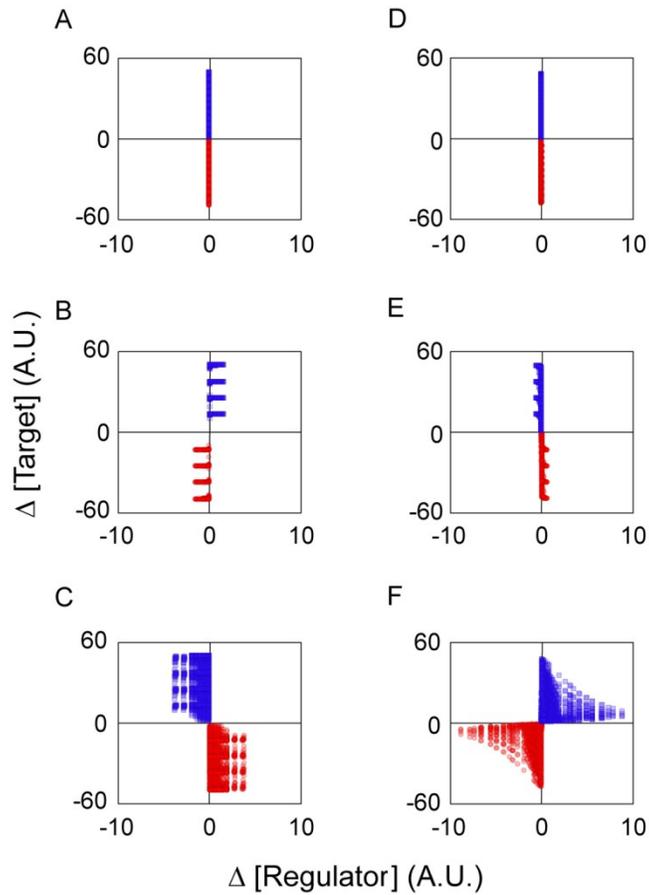


Figure S7. Change in total regulators and target in switching the topology ON and OFF. Change in steady state target response (ΔTarget) in switching ON (blue dots) and OFF (red dots) against change in amount of regulators ($\Delta\text{Regulator}$) have been plotted for networks in “Performance Box”. Regulators include both active and inactive form.

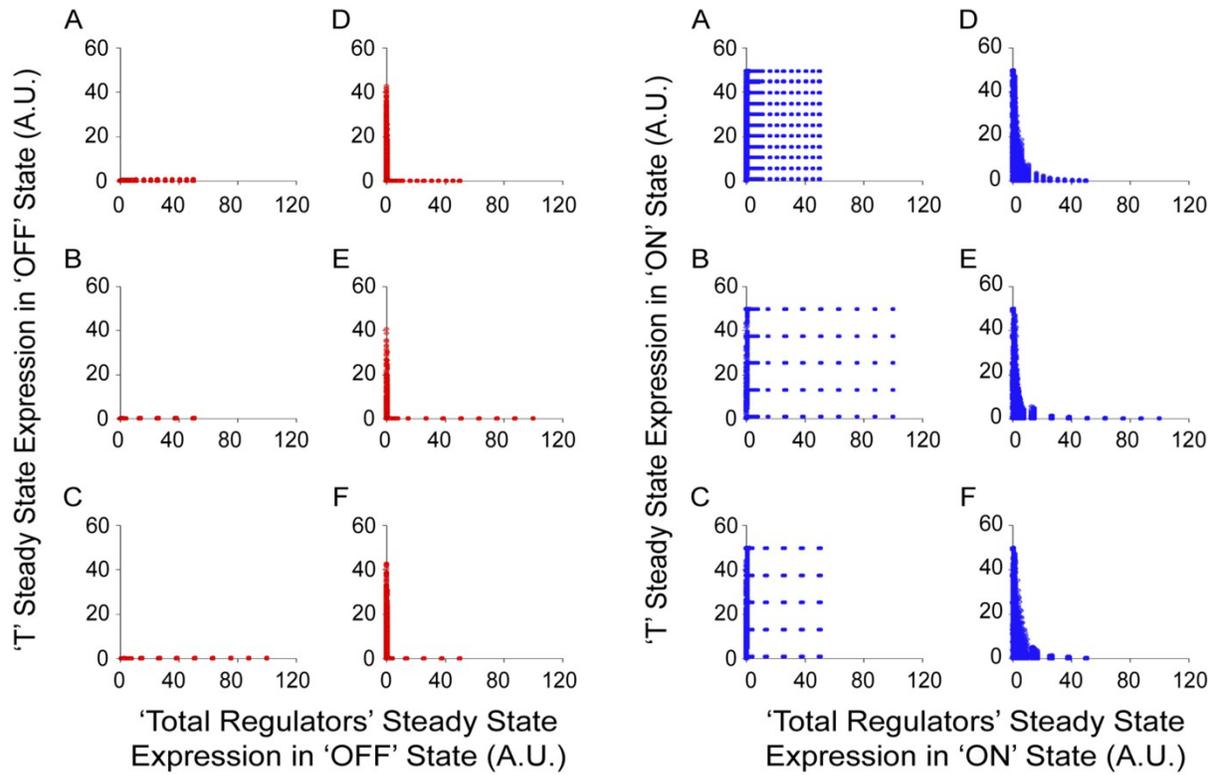


Figure S8. Target response level against total regulators at steady state.All networks (~90000) were included in this plot. Blue and red color indicates ON and OFF conditions respectively.

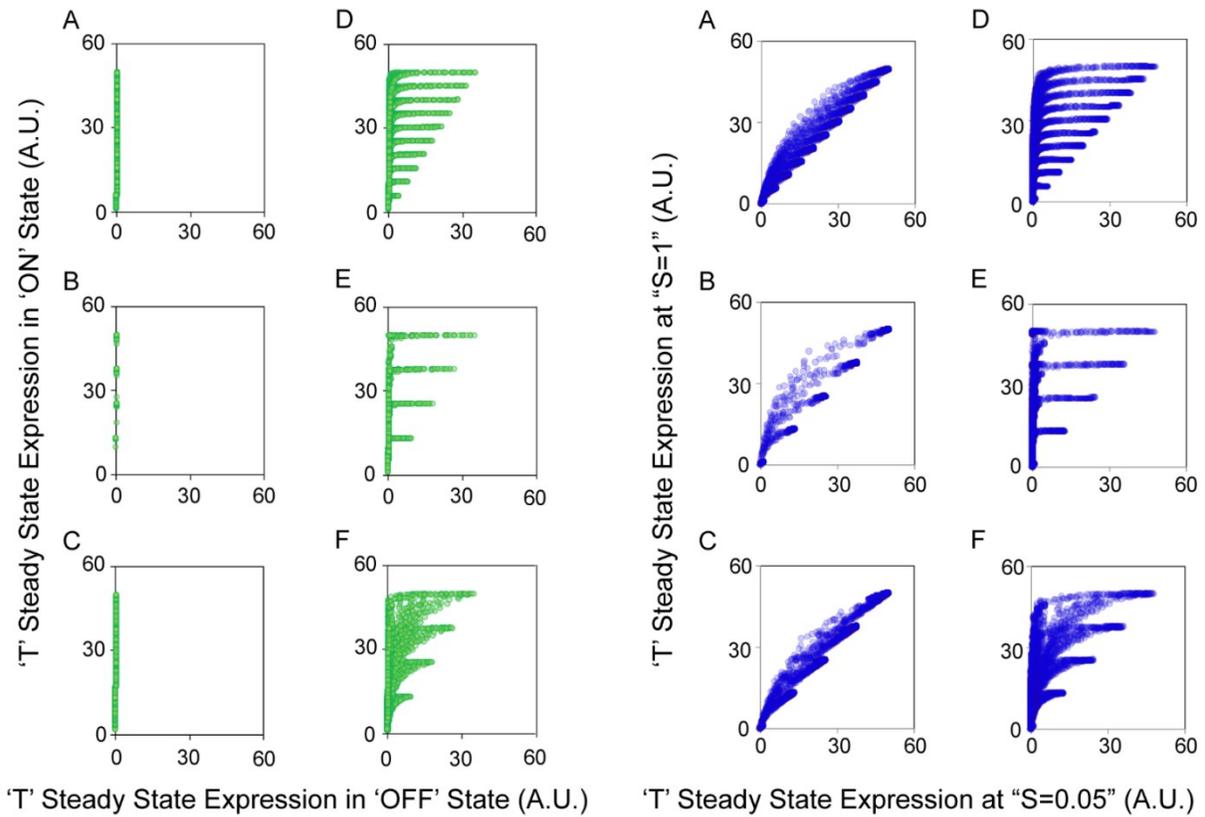


Figure S9. "Switchability" and "Sensitivity" of networks. (Left Panel) Steady state target T in the OFF (x-axis) vs. the ON condition (y-axis) for all six topologies in "Performance Box". **(Right Panel)** Steady state target T for all networks (both outside and inside "Performance Box") at very low and very high signal strength were calculated and plotted.

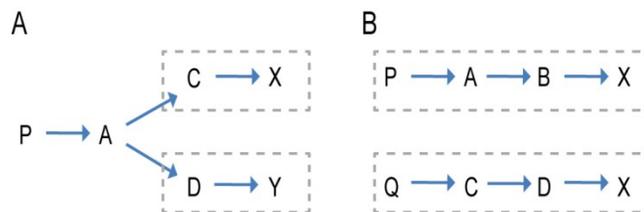


Figure S10. (A) A hypothetical example showing the steps involved in biosynthesis of an amino acid. The steps shown in upper and lower boxes are unique to synthesis of amino acid X and Y respectively. **(B)** A hypothetical example showing steps when an amino acid is synthesized by two different pathways.