

Modelling Carbon Fixation and G-encoded Amino Acids

Method

The Gillespie algorithm was employed for creating stochastic simulations through the GillespieSSA package on R. First, models were created based on modern metabolism through reactions identified in MetaCyc. Then, equivalent simplified models were created, where the need for modern cofactors was eliminated. The direct method was used, which sampled one reaction at a time to project a continuous trajectory. In order to do this, an initial state vector, state-change matrix, and propensity vector were supplied.

The initial state vector contained concentrations for all the reactants and products involved in the reactions considered. Arbitrary values were assigned, with 5000 for reactants and cofactors and 0 for all products, in a cell volume of 1. The state-change matrix had stoichiometric ratios for the reactants and products in every reaction, with rows corresponding to species and columns to reactions. The propensity vector contained rate equations for every reaction step. The reactions were classified based on EC number and rate constants were initially assigned based on this, where EC categories most commonly seen in the pathways were assigned higher rate constant values. Given this, the algorithm characterized each reaction with a propensity function and state change vector.

The algorithm sampled two numbers to create a projected trajectory. One number was sampled from the distribution given by the sum of propensity functions in order to determine the time to the next reaction. The second number was sampled from a probability function from the propensity vector to determine which reaction fired next. The trajectory was then generated by starting from the initial state and firing reactions in continuous time, updating the system state at every time step.

Reactions were modelled kinetically, where the algorithm considered the rate equations as proportional to the probability of a reaction occurring. As such, higher rate constants and higher order reactions indicated a greater probability of the reaction proceeding, corresponding to faster, more kinetically favorable reactions. In the future, these models can be extended to include thermodynamic parameters. Reactions were also modelled in mass action.

All code and data used in the models can be found at https://github.com/liabote/summer_proj.

Preliminary Results

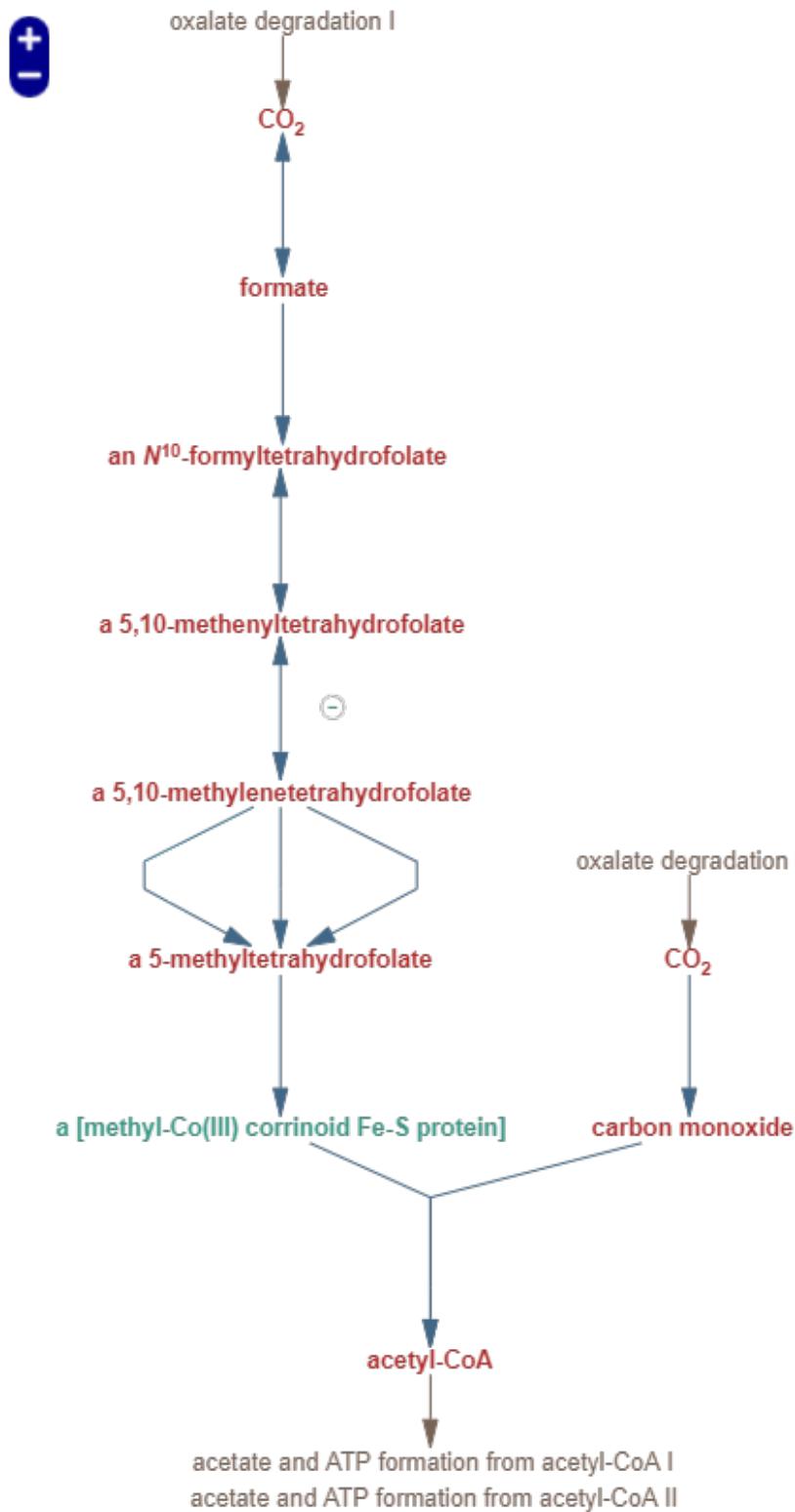
G-encoded Amino Acids

Given this, the initial model was reduced to only include carbon fixation, the earliest pathway in autotrophic metabolism, and glycine synthesis, which immediately follows it. This was done to inspect if earlier intermediates were being produced and consumed consistently with the reactions considered.

The Ljungdahl-Wood pathway was used to model carbon fixation. This pathway occurs in autotrophs, and is best understood in homoacetogenic bacteria, to result in the production of acetyl-coA from two molecules of CO₂. Understanding the carbon fixation pathway is key to modelling prebiotic metabolism, as carbon fixation allows for the establishment of autotrophic growth and development by reducing dependence on organic material for energy.

This resulted in lower levels of glycine synthesis than in the model which included all the amino acids (Appendix, Figure 2). Acetyl-coA was not produced at the end of carbon fixation. Most earlier intermediates

were also not produced, or produced only at low concentrations, except for carbon monoxide. As such, the carbon fixation pathway did not occur completely. In order to best explore the reasons behind this, the individual reactions in the carbon fixation pathway were further analyzed.



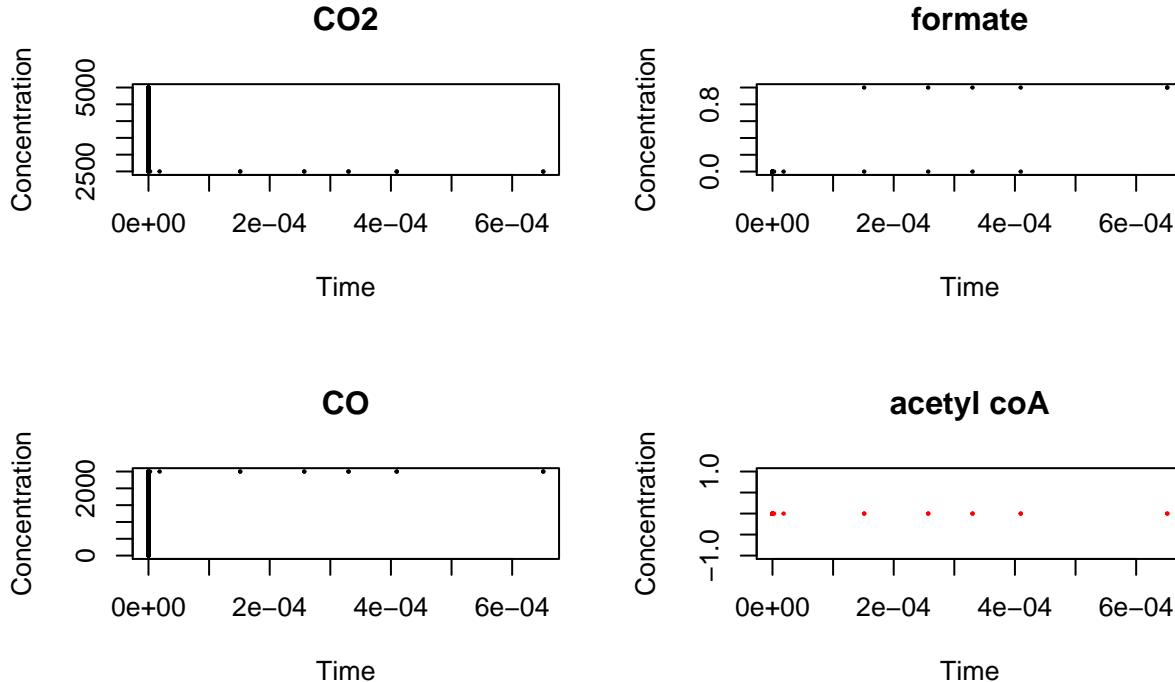
Carbon fixation

As shown in the figure above, CO_2 is used in two parallel branches for carbon fixation. The first involves six steps from the conversion of CO_2 to methyl corrinoid (III) FeS clusters. The second involves a single step converting CO_2 to CO. Methyl corrinoid (III) FeS and CO are then used as inputs in the final step of CO_2 fixation, acetyl coA synthesis. The carbon fixation model showed high levels of CO synthesis. However, methyl corrinoid (III) FeS was not produced. The tetrahydrofolates that are meant to be produced in preceding steps were also not synthesized, with the exception of N10-tetrahydrofolate, which is the earliest one produced in this pathway.

This may be explained by competition for reactants either between the two parallel pathways, or within the pathway for methyl corrinoid (III) FeS synthesis. NADPH is consumed when CO_2 is used for formate synthesis, and again later on for the synthesis of 5,10-methylenetetrahydrofolate from 5,10-methyltetrahydrofolate. Formate is produced in oscillating concentrations and used in the synthesis of N10-tetrahydrofolate, which unlike 5,10-methylenetetrahydrofolate is successfully produced in this model. As such, NADPH may be consumed early in the pathway and depleted before 5,10-methyltetrahydrofolate can be produced. Alternatively, CO production from CO_2 involves only one step, which is a fifth order reaction and thus relatively rapid compared to formate synthesis that is only second order. CO_2 may be consumed more rapidly in CO synthesis than in the multi-step pathway to methyl corrinoid (III) FeS production, resulting in the high levels of CO produced in the model.

To test these hypotheses, a model was created including only the reactions for carbon fixation, introducing a few modifications.

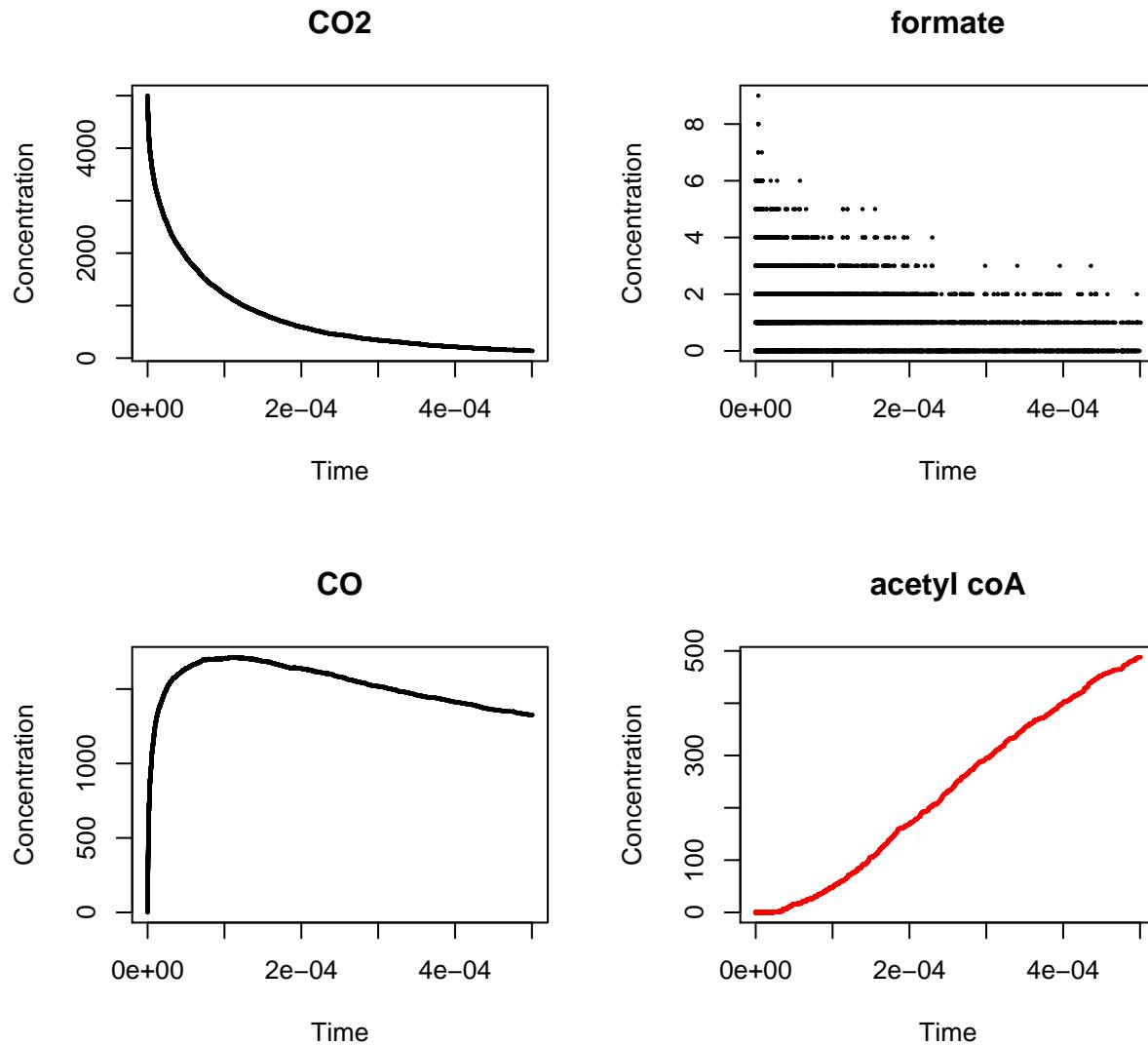
Removing the requirement for NADPH in formate synthesis



In the figure above, the NADPH requirement was removed from formate synthesis, such that NADPH was only consumed for the production of 5,10-methylenetetrahydrofolate. However, while no NADPH depletion was seen in the model and acetyl-coA was not produced, CO_2 was still rapidly consumed. This suggests that the pathway for methyl corrinoid (III) FeS synthesis was not occurring, and that the CO_2 was funneled into

CO synthesis. The internal competition for NADPH amongst tetrahydrofolate intermediates must not be a strong rate-limiting step.

Decreasing the rate of reaction for CO synthesis

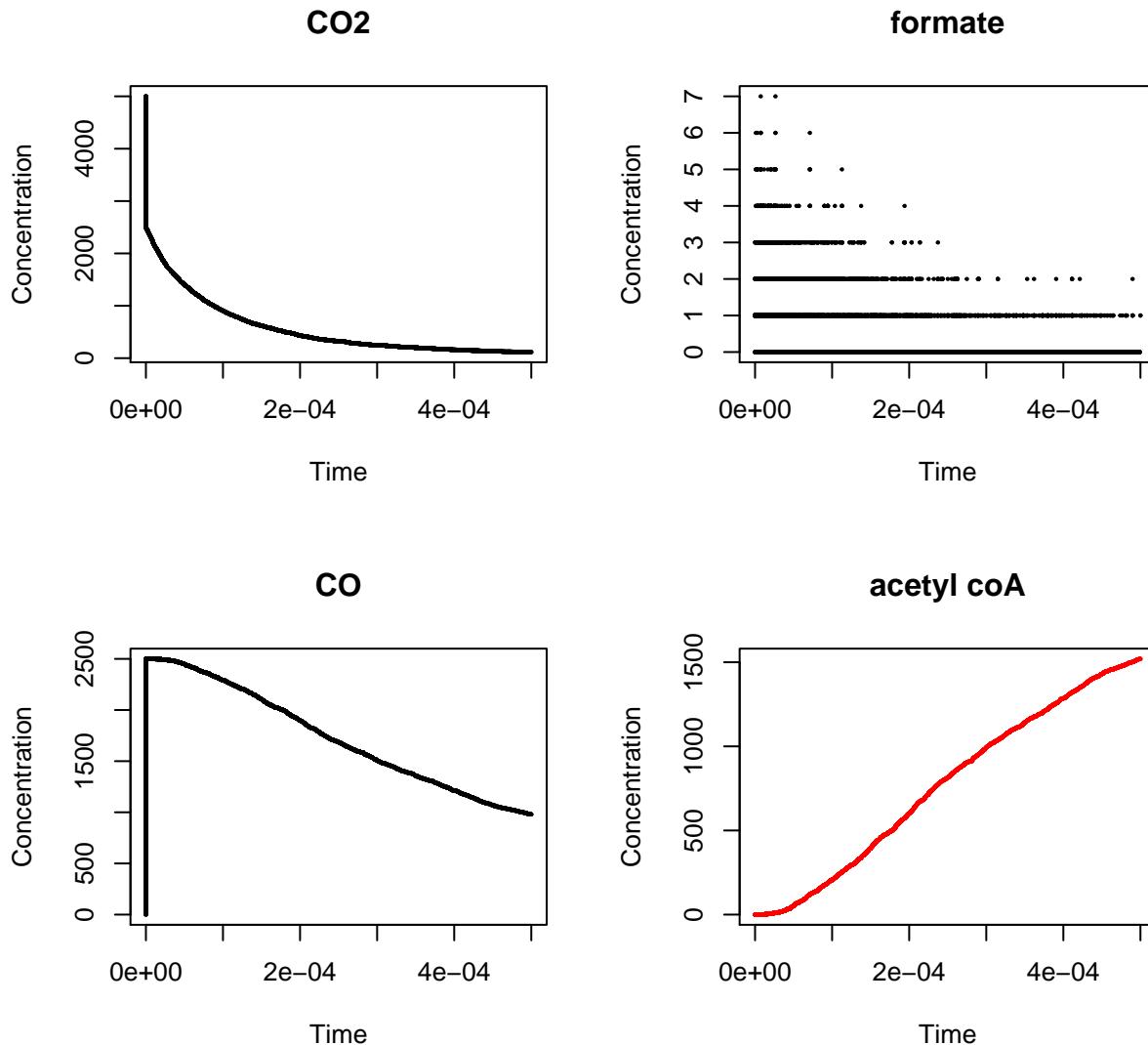


The figure above shows the results of a model where the rate of CO synthesis is decreased by a factor of 10^{10} . Production of all the tetrahydrofolate intermediates occurred, although at a lower level for intermediates synthesized later in the pathway. CO and low levels of methyl corrinoid (III) FeS were both synthesized, which reacted to produce acetyl coA. This suggests that CO synthesis is likely the limiting step for this pathway. These results are seen when CO synthesis is decreased by at least a factor of 10^3 .

Making CO synthesis second order like formate synthesis, as opposed to the original fifth order, also generates the same results. This indicates that when CO and formate synthesis reactions compete for CO_2 at the same rate, carbon fixation proceeds to completion. Competition for CO_2 between the parallel pathways is thus rate-limiting.

In subsequent models, the reactions were set to first order with respect to each species. This is due to the lack of kinetic data which can inform us about appropriate rate orders.

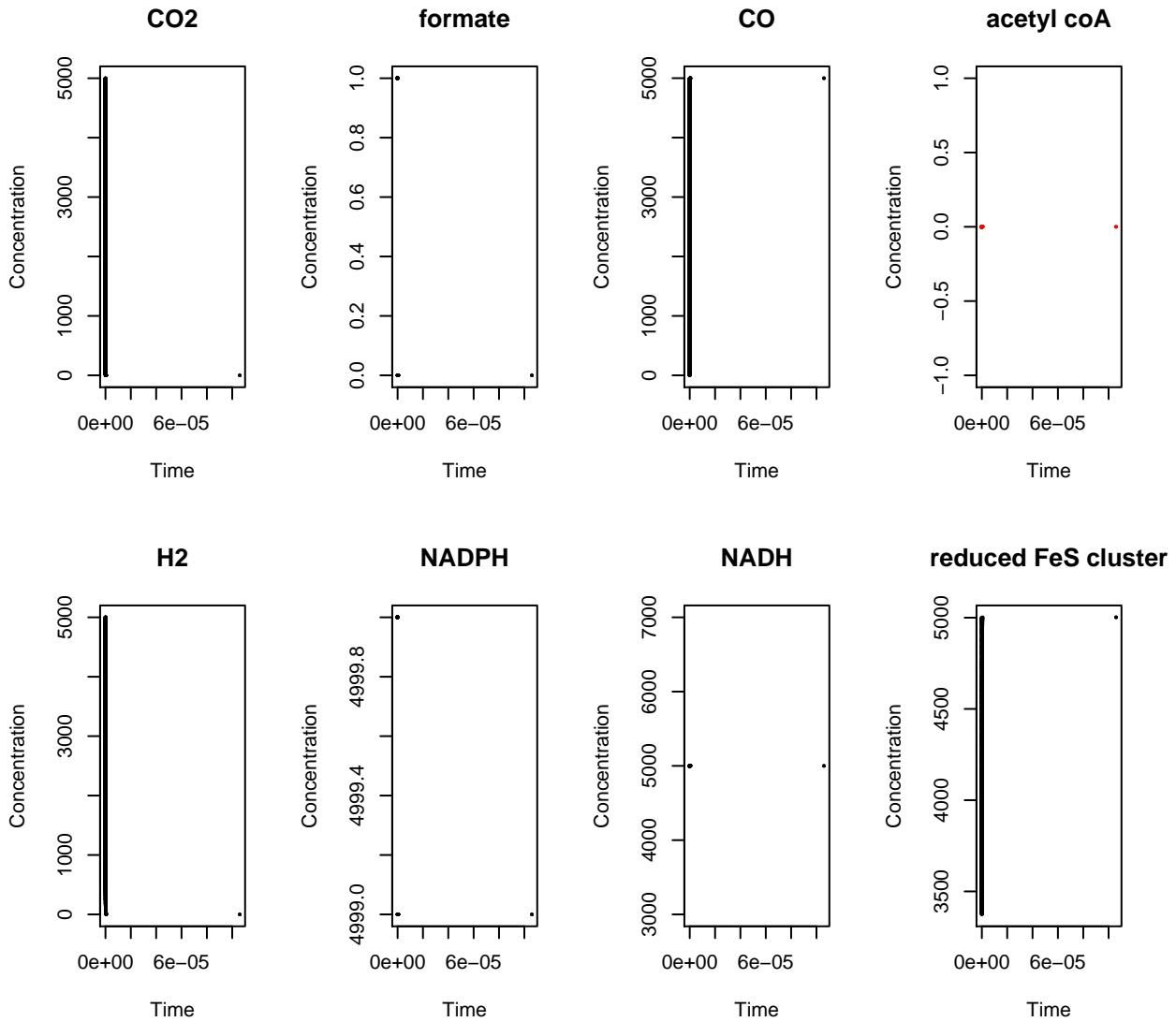
Increasing the initial H⁺ concentration



In the figure above, the initial H⁺ concentration was increased by a factor of 3. This similarly results in the synthesis of CO, all tetrahydrofolate intermediates, and acetyl coA. This may be because both CO synthesis and many of the steps to methyl corrinoid (III) FeS production involve redox reactions, but most of the available protons are consumed for the faster reaction, CO synthesis. Increasing H⁺ concentration prevents inhibition of the pathway by H⁺ depletion, allowing methyl corrinoid (III) FeS synthesis to proceed.

Regenerating reducing agents

The competition for reagents and reducing agents suggests that a model with only forward reactions, unable to regenerate the reactants consumed, will inhibit itself over time. Reducing agents will be regenerated after they are oxidized if they are able to react with hydrogen in the environment. The introduction of reactions for NADPH, NADP, and reduced FeS cluster synthesis is modeled below.



Reduced FeS clusters were generated, although acetyl coA, NADPH, and NADP were not. This may be because the CO synthesis pathway still occurs at a greater rate, such that most of the available hydrogen is being used to regenerate only the reduced FeS clusters that are consumed in CO synthesis. Reduced FeS cluster synthesis is also third order, while NADPH and NADH are only second order overall, which means it is occurring at a faster rate and able to sequester more hydrogen than the other two regenerative steps.

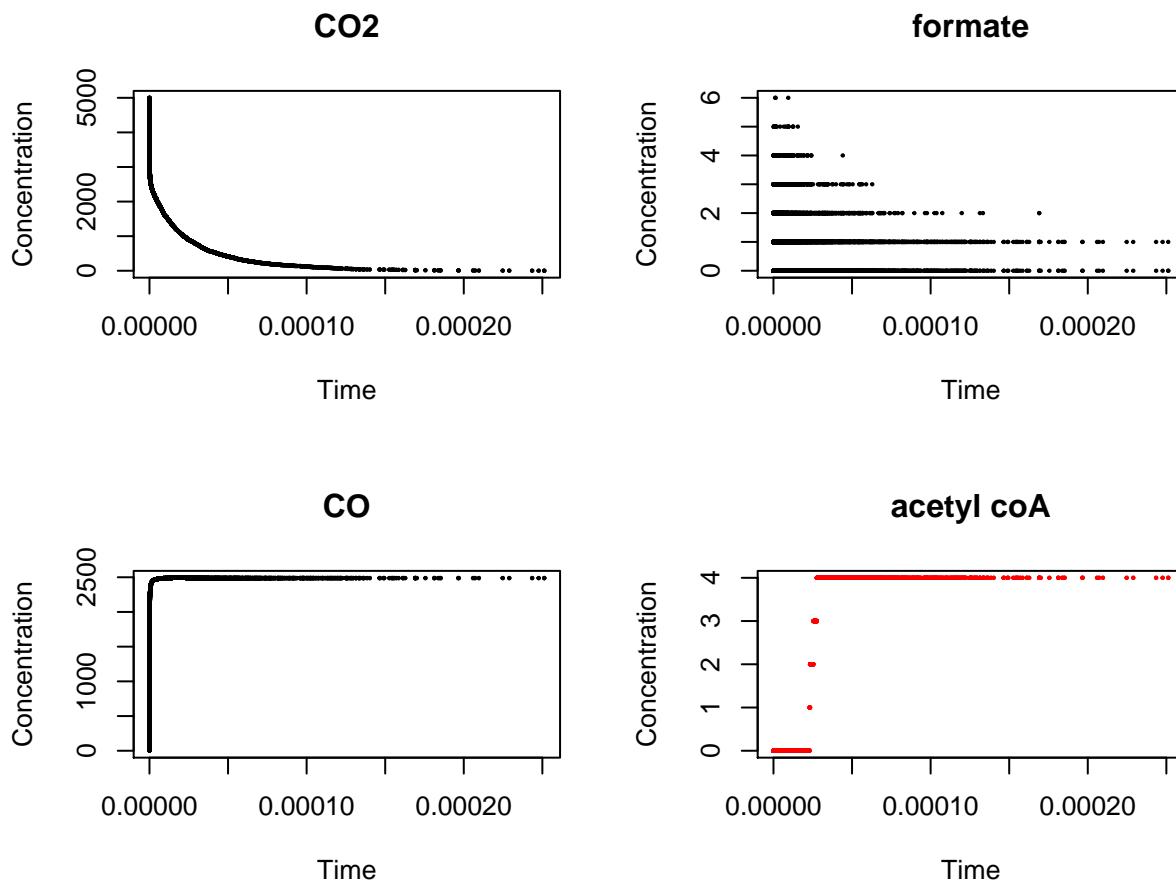
Changing the rate constants

Rate constants were obtained for every reaction in the pathway based on the reported k_{cat} of each, although values were only available for enzyme-catalyzed reactions based on experimental models in physiological conditions. k_{cat} values were chosen from enzymes found in bacteria or archaea, for reactions going in the direction that they occur in metabolism. These k_{cat} values are listed in Table 1 below, with reactions listed in the order they appear in Figure 1.

Table 1. Enzymes in Carbon Fixation and Corresponding K_{cat} (s^{-1})

Reaction	K_{cat} (s^{-1})	Enzyme
1	10	formate dehydrogenase
2	4.4	formyltetrahydrofolate ligase
3	16	bacterial methenyltetrahydrofolate cyclohydrolase
4	19.9	methylenetetrahydrofolate dehydrogenase
5	23.5	methylenetetrahydrofolate reductase
6	108	5-methyltetrahydrofolate: corrinoid/ FeS protein Co-methyltransferase
7	0.869	CO dehydrogenase/acetyl-CoA synthase
8	0.061	CO dehydrogenase/acetyl-CoA synthase

These values were then supplied as initial parameters in the model for carbon fixation, and the results of this are shown below.

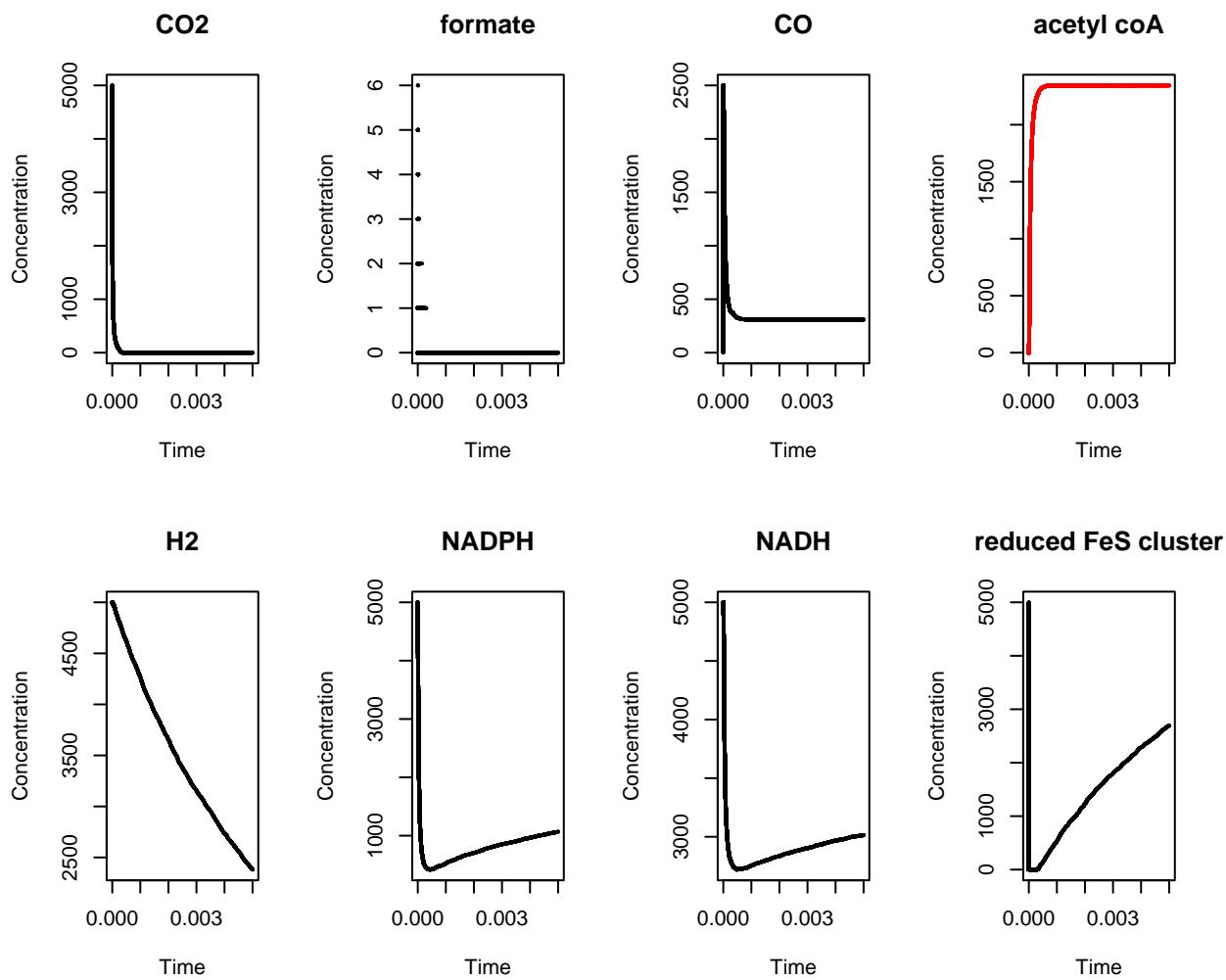


Low levels of acetyl-coA synthesis occurred, which was not detected in the unmodified model. CO was still successfully being produced at high concentrations. Formate and proceeding tetrahydrofolate intermediates

were synthesized successfully, although only in very low or oscillating concentrations. Formate synthesis has a k_{cat} two orders of magnitude greater than that of CO synthesis in physiological conditions, thus allowing for the methyl corrinoid (III) FeS synthesis pathway to proceed despite the problems described earlier.

Combined model

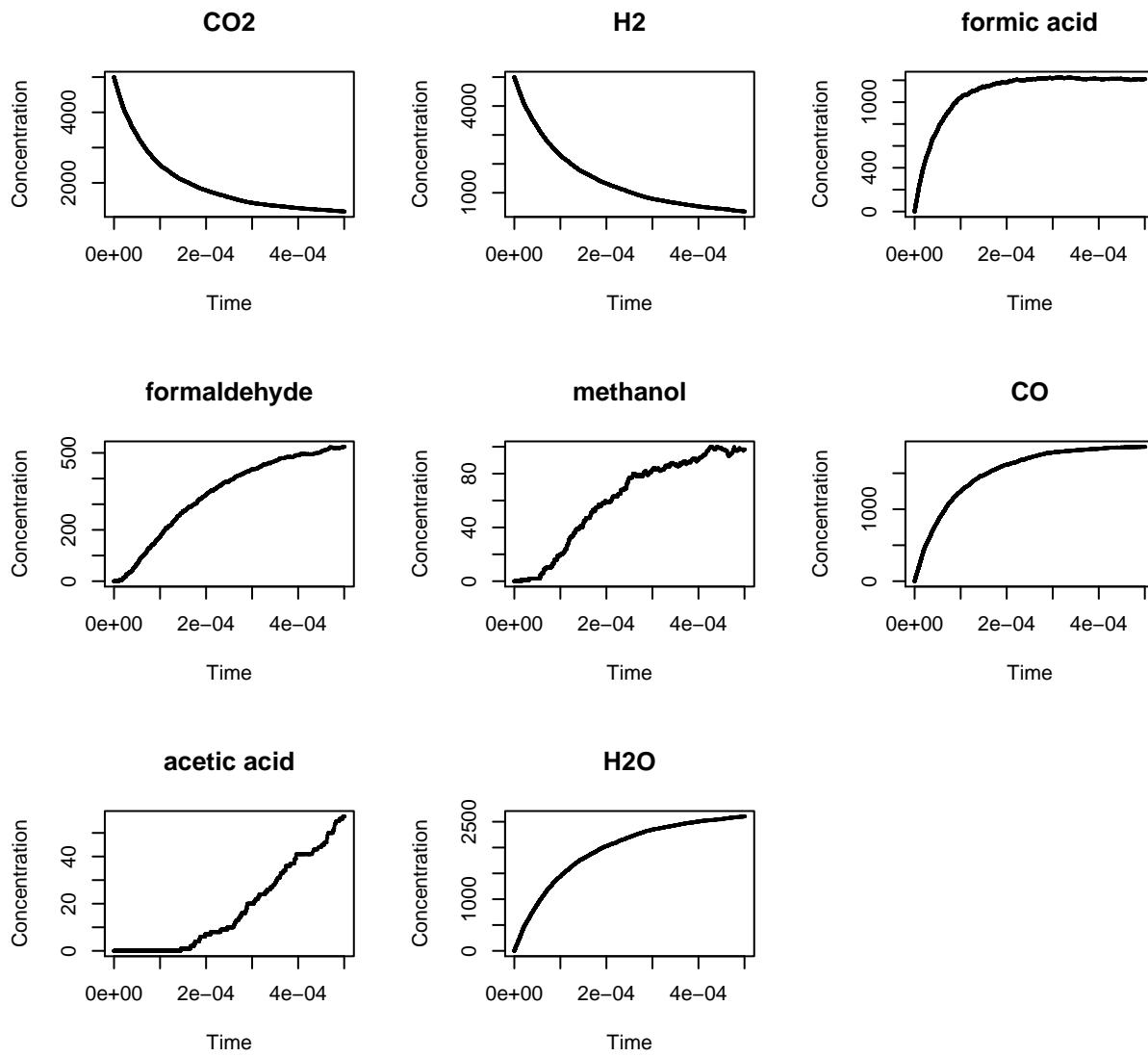
A model was then created with these rate constants from literature, an excess of H^+ to simulate a reducing environment, and reactions for the regeneration of reducing agents. The regenerative reactions were given a reaction rate one order of magnitude less than the slowest forward reaction. The reaction for reduced FeS synthesis was also made first order with respect to oxidized FeS and thus second order overall. The results of this are shown below.



Acetyl-coA, formate, and CO were all produced. The tetrahydrofolate intermediates were also synthesized, alongside all the reducing agents, although reduced FeS clusters were still regenerated at a faster rate than NADPH and NADH. While the modifications introduced resulted in the successful synthesis of the expected products of carbon fixation, other conditions and rates of reactions may still be explored. An excess H^+ gradient may not be the only condition for successful acetyl-coA synthesis.

Removing cofactors

A simpler model was created without the need for cofactors using rate constants from literature. Only CO₂, H₂, and NH₃ were initially supplied. The results of this are seen below. CO₂ is successfully fixed to acetic acid. CO synthesis is less kinetically favorable, with a significantly lower rate constant, than methanol synthesis and all preceding steps. Thus, more methanol than CO was produced, in contrast to earlier models where CO synthesis was higher. However, when all reactions were assigned a rate constant of 1, more methanol was produced than CO. Thus, the relative amounts of each product synthesized were dependent on the initial parameters supplied. The results of this are shown below.



First five amino acids with no cofactors

This model was then extended to include the G-encoded amino acids glycine, glutamate, alanine, valine, and aspartate. When each reaction was given a rate constant of 1, or assigned its rate constant from literature, glycine was the only amino acid produced. This is because glycine used all of the initial inputs supplied and had the greatest value for its rate equation, where other amino acids only used NH₃ and H₂.

The amounts and ratios of CO₂, H₂, and NH₃ supplied were then changed to maximize the number of products synthesized. Increasing CO₂ and H₂ was found to result in more products being synthesized as these inputs are required in all the pathways, following the first step of carbon fixation between CO₂ and H₂. H₂ was used in more reactions and was increased by a multiple of 5, whereas CO₂ was increased by a multiple of 3. Increasing CO₂ by a higher multiple favored synthesis of glycine but not other products or amino acids. Conversely, reducing NH₃ resulted in more products synthesized, as only amino acid synthesis used NH₃ as an input, and setting a very high initial NH₃ concentration drastically favored the glycine synthesis step. Changing these initial concentrations resulted in the production of acetic acid and pyruvic acid, indicating completion of carbon fixation.

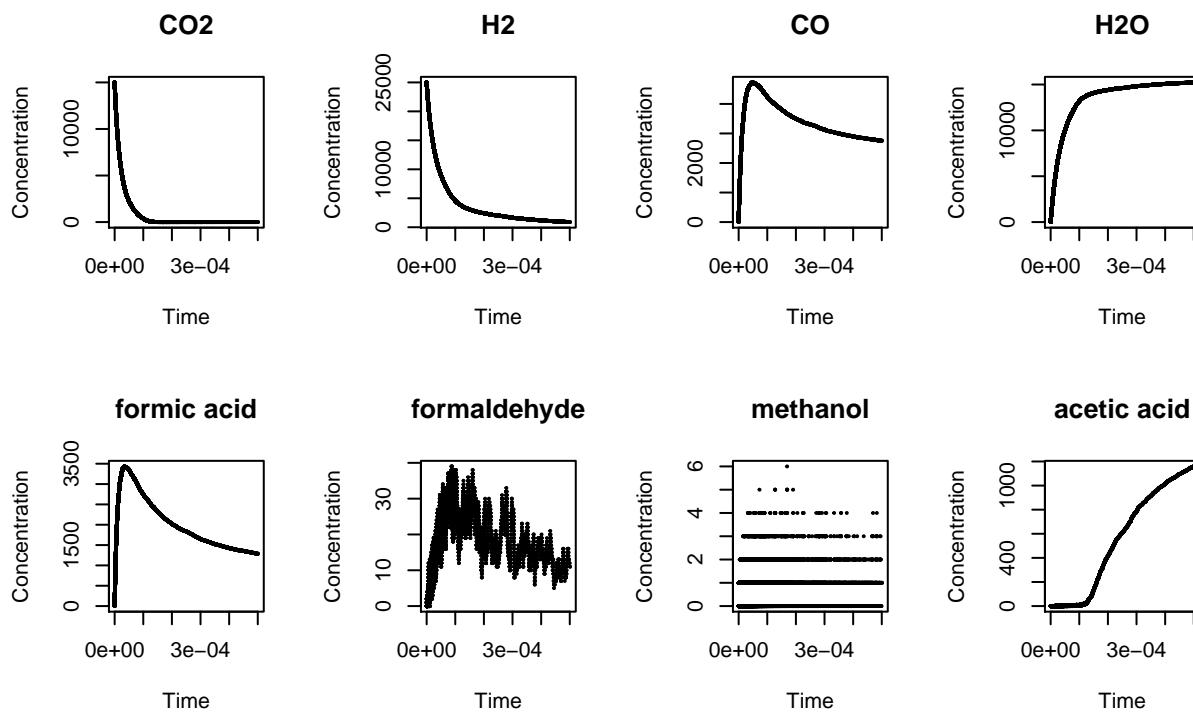
The reactions were then assigned classes based on the reaction type. However, assigning a rate constant to each class, based on the average values from literature, did not improve the model.

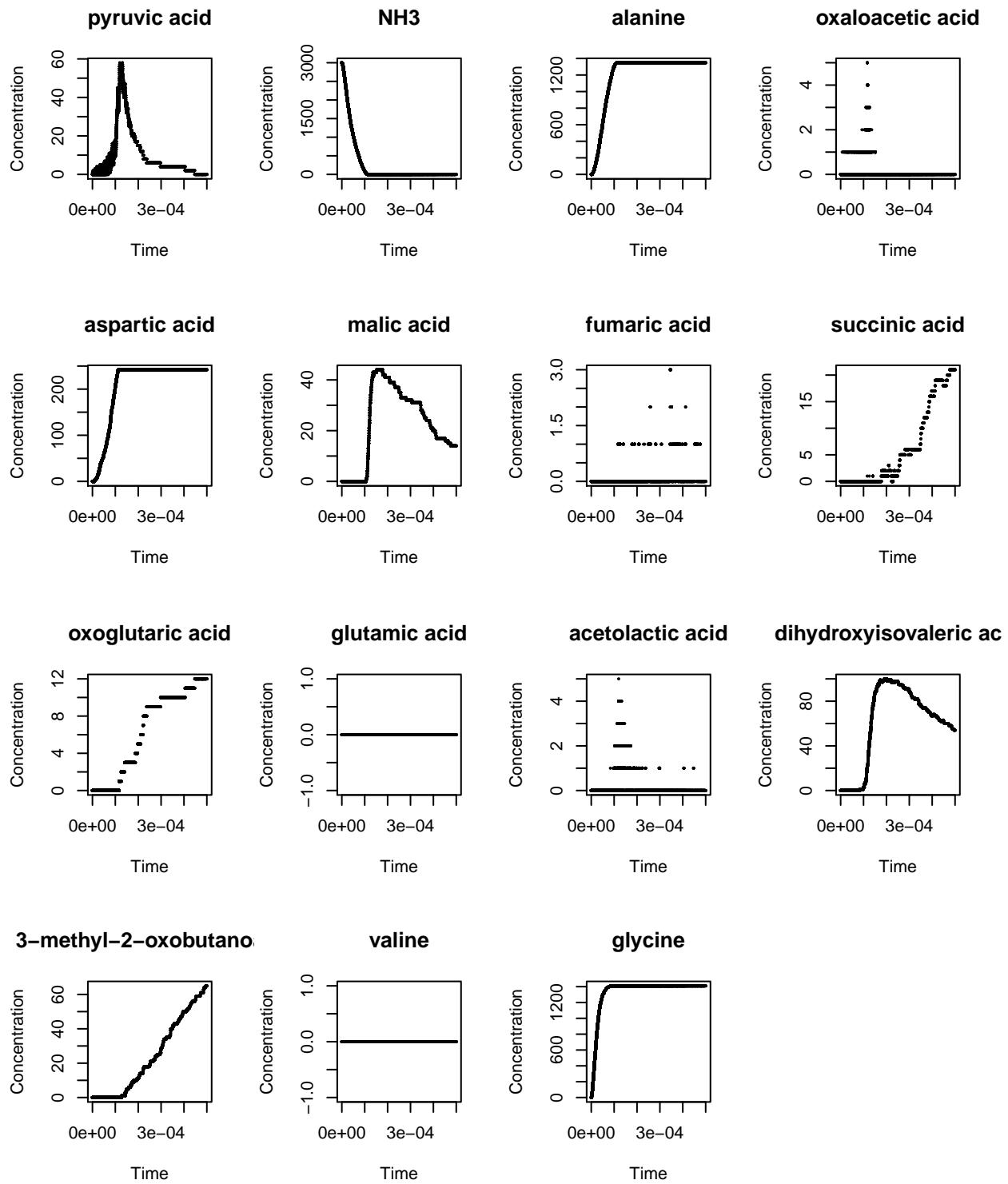
Synthesis of all intermediates and amino acids was only achieved by changing individual rate constants, although glutamic acid and valine were sometimes produced in very low concentrations or not at all. The results of these changes are shown in the figure below. First, the quickest reactions were slowed down by dividing them by a factor, and the slowest reactions were sped up by multiplying them by a factor. For example, reducing the rate of glycine synthesis allowed for low levels of alanine synthesis.

Then, rates of reactions competing for the same inputs that are synthesized as intermediates in the pathways were modified. Pyruvic acid, for instance, is used for alanine, oxaloacetic acid, and acetolactic acid synthesis. However, alanine synthesis is the fastest of these reactions, and was thus using up all the pyruvic acid produced. Multiplying the oxaloacetic acid and acetolactic acid rate reactions by a factor allowed the pathways to proceed.

Finally, the rate of each step was limited by the preceding step. Reactions that had a significantly greater rate than its preceding step were not successfully completed, as enough inputs could not accumulate. As such, some reaction rates were adjusted accordingly.

Rate constants were also assigned to classes based on the average relative rates of each class, although successful synthesis of all amino acids was not achieved. Additional parameters, such as thermodynamic constraints, may be required in future models to more accurately design and understand the models.





Appendix

Figure 1. Models for the synthesis of G-encoded amino acids

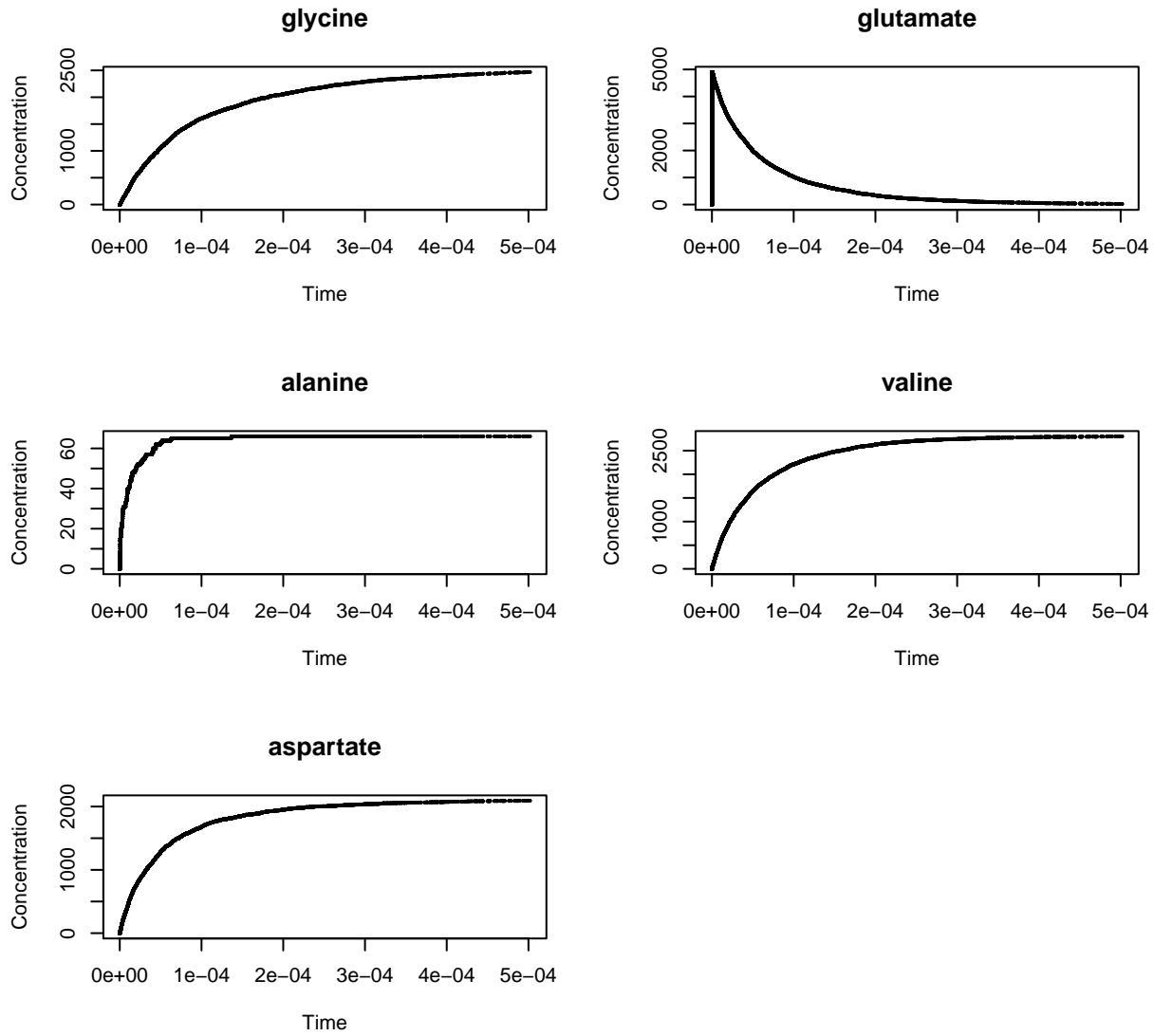


Figure 2. Model from carbon fixation to glycine synthesis

