We thank the reviewers for their helpful suggestions to improve the manuscript, "Dynamic Modeling of Cell-Free Biochemical Networks using Effective Kinetic Models." Below, we have outlined the changes we made to the manuscript to address reviewer suggestions.

Reviewer 1

We respectfully disagree with the critique presented by reviewer 1. Given that many of the comments made by reviewer 1 were diametrically opposed by the other reviewers, we have not exhaustively addressed the comments of reviewer 1. However, the one exception was the issue of experimental design. Experimental design was raised by the other two reviewers as well. To this end, we have updated the discussion section, see point 1 for reviewer 2 and point 8 for reviewer 3.

Reviewer 2 (Changes highlighted in Yellow on MARKED manuscript)

1. As discussed by the authors (Figures 7 and 8), their method may not be effective in identifying the correct control structures. This issue may be taken care of to an extent when the network size is moderate (by designing new experiments and collecting addition data), but how can this problem be overcome for “large” networks?

   We completely agree with the reviewer that identification of allosteric connectivity will be a critical challenge moving forward. Toward this concern, we have included in the discussion section a description of recent work by Sauer and colleagues that may be a useful technique for model-guided identification of cell-free allosteric control networks:

   “Recently, Sauer and colleagues have developed a systematic, model-based approach for the identification of allosteric regulation *in vivo*. They tested the effects of many putative allosteric protein-metabolite interactions on the performance of a kinetic model of glycolysis against dynamic metabolomic and fluxomic measurements. A method similar to this may be easily applied to cell-free systems in order to identify relevant *in vitro* allosteric interactions. Because omics measurements of cell-free environments are easy to obtain, identification of large-scale allosteric control structures may be possible.”

   New Reference:

   Link H, Kochanowski K, Sauer U (2013) Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. Nat Biotechnol 31: 357-61.

2. The reviewer wonders if the proposed method can be used for simulating actual cells (by modifying enzyme balance equations and adding the growth term) as well as cell-free networks.

   We thank the reviewer for this suggestion. Indeed, our approach can be extended for the simulation of cell-based systems. In order to highlight this advantage, the following text has been added to the end of the second paragraph of the Discussion:

"Though requiring a more complex description of cellular metabolism, our approach may even be extended to simulate cell-based systems by incorporating the same rule-based laws into transcription factor activation and gene expression regulation."

3. As the proposed method may result in a set of candidate models, there is an analogy with "ensemble modeling" by the James Liao group (Linh M. Tran Matthew L. Rizk James C. Liao, 2008, Ensemble Modeling of Metabolic Networks, Biophysical Journal, 95(12): 5606-5617). Discussion on the conceptual similarity and difference between two approaches would be helpful.

   We thank the reviewer for this useful suggestion to improve the manuscript. We've included the following text in the last paragraph of the Discussion:

   "Ensemble modeling is a well-established approach for parameter identification in large-scale deterministic models. Liao and coworkers developed a method that generates an ensemble of kinetic models that all approach the same steady-state, one determined by fluxomics measurements. The best subpopulation of candidate models are selected based on their agreement with further measurements of genetically perturbed systems. Our work relies on heuristic search optimization to identify kinetic models consistent with, typically, dynamic time-series measurements of cellular species. Instead of estimating a single yet highly uncertain parameter set, both approaches estimate an ensemble of parameter sets whose model behavior recapitulates experimental measurements."

   New Reference:

   Tran LM, Rizk ML, Liao JC (2008) Ensemble modeling of metabolic networks. Biophys J 95: 5606-17.

4. The reviewer does not agree with some of the statements on the cybernetic approach: The first paragraph on page 2: "Unfortunately, cybernetic models also suffer from an identifiability challenge…."– This would be true for fully structured cybernetic models such as the ones used by Young et al. (Biotechnol Bioeng, 100(3):542-59, 2008) or Varner and Ramkrishna (Metab Eng, 1(1):88-116, 1999), but is not an issue in the recent formulations called hybrid cybernetic models (HCMs) based on elementary mode reduction (Song and Ramkrishna, Biotechnol Bioeng, 102 (2): 554-568, 2009; Song et al., Biotechnol Bioeng, 103 (5): 984-1002, 2009) and lumped HCMs (L-HCMs) (Song and Ramkrishna, Biotechnol Bioeng, 106 (2): 271-284 , 2010; Biotechnol Bioeng, 108 (1): 127-140, 2011). The authors should remove or consider appropriately revising that statement.

   We thank the reviewer for this insightful comment. We agree that there exists cybernetic based approaches that successfully deal with large-scale parameter identifiability challenges. Here we were referring to detailed structured cybernetic models. The text at the end of the first paragraph of the introduction has been altered to reflect this as well as include the alternative cybernetic methods for model reduction as follows:

   "Unfortunately, traditional, fully structured cybernetic models also suffer from an identifiability challenge, as both the kinetic parameters and an abstracted model of cellular objectives must be estimated simultaneously. However, recent cybernetic formulations from Ramkrishna and colleagues have successfully treated this identifiability challenge through elementary mode

reduction, though the techniques replace detailed biological mechanism with an optimization heuristic.

New Reference:

Song HS, Morgan JA, Ramkrishna D (2009) Systematic development of hybrid cybernetic models: application to recombinant yeast co-consuming glucose and xylose. Biotechnol Bioeng 103: 984-1002.

5. The second paragraph on page 14: “However, cybernetic control heuristics are not mechanistic ...”–The opinion viewing the cybernetic control laws as heuristic or non-mechanistic needs to be changed because their theoretical foundations were established based on optimal control theories by Young and Ramkrishna (Biotechnol Progress, 23(1): 83-99, 2007).

We thank the reviewer for this comment. However, we respectfully disagree that cybernetic control laws are biologically mechanistic. We agree that cybernetic control laws have a solid basis in optimal control theory. However, this does not mean they are mechanistic in a biological sense. Cybernetic control laws are optimization heuristics that replace biologically mechanistic descriptions of gene expression and enzyme activity regulation. That is their power. However, they do not provide insight into specific mechanistic interactions, instead they abstract them. Thus, we do not feel the language here needs to be altered.

6. It is unclear what “N” means throughout the manuscript.

For clarity, all instances of “N” have been removed from the manuscript text. The “N” in Figure 3 has been replaced with η to be consistent with Equation 10. Also, in Figure 3, we replaced $k_{control}$ with $\kappa_{control}$ to be consistent with Equation 10.

7) Implications of eta in Eq. (10) need to be described

In order to clarify the meaning and implication of cooperativity control parameter η, the following text has been added to the Materials and Methods sections after Equation 10:

“In the case $\eta > 1$, the allosteric interaction displays positive cooperativity. For $\eta < 1$, the interaction is negatively cooperative. Finally, if $\eta = 1$, the interaction displays no cooperativity. The effect of different values of η on reaction rate can be seen in Figure 3.”

7. Last line of page 16: the reference for LSODE routine should be provided.

The Octave and LSODE reference is now provided in the text as (v 3.8.1; www.octave.org).

Reviewer 3 (Changes highlighted in Green on MARKED manuscript)

1. It is not clear how Figure 3 was generated from the framework. What control parameter values were used. Were these parameter taken as a random sample? Also, wouldn't the behavior shown in Figure 3 be expected from the transfer function (Eq. 10) used in the model?

   We thank the reviewer for this comment. Fig 3 presents a simple proof of concept application of the framework to model well understood enzyme kinetic expressions. The results show that we can capture the expected cooperativity or inhibition behavior without mechanistic knowledge. Additionally, it shows that we can modify standard kinetic expression such as multiple saturation kinetics with transfer functions to capture cooperativity that was not present in the original expression.

   We have updated the caption to reflect the control and cooperativity values used in the simulations.

2. Related to Figure 3, the author stated that "increased substrate abundance decreased the reaction rate". Where has this been shown? In both Figure 3(A), the reaction rate increased with substrate for all N considered.

   We thank the reviewers for this helpful comment. We meant to say that increased substrate level decreased the maximum rate for cooperativity values less than unity. We have corrected the manuscript to address this error. The new text reads:

   "… For cooperativity parameters less than unity, increased substrate abundance decreased the maximum reaction rate. "

3. Have the authors considered other implementations of the rules than shown in Figure 2? Instead of using max or min, would a multiplication of possible f_i(Z)s also work? The use of min/max function would introduce a discontinuity around certain time points. How was such discontinuity handled during the integration of the ODE model?

   We thank the reviewer for this excellent question. We have not considered additional integration rules or different transfer function models in this study. However, we expect potentially interesting properties could arise from different choices of these factors. The reviewer also raised an important point about potential discontinuities arising from the max/min formulation of the integration rules. The max/min formulation, which is similar to cybernetic variables, did not cause numerical instability in this study, nor in previous cybernetic modeling studies conducted by the senior author. Alternative integration rules such as the mean might have different properties which could influence model identification or performance. For example, a mean integration rule would be differentiable, which allows derivative-based optimization approaches to be used. The particular form of the transfer function could also be explored. We choose a Hill-like function because of its prominence in the systems and synthetic biology community. However, the only mathematical requirement for a transfer function is that it map a non-negative continuous or categorical variable into the range [0,1]. Thus, many types of transfer functions are possible. In summary, while we have not explored this question in the current study, we have several on going research projects where the choice of integration rule and transfer functions are being explored. Toward this point, we have added the following text the last paragraph of the discussion:

"… Lastly, the choice of max/min integration rules or the particular form of the transfer functions could be generalized to include other rule types and functions. Theoretically, an integration rule is a function whose domain is a set of transfer function inputs, and whose range is $v\in[0,1]$. Thus, integration rules other than max/min could be used, such as the mean or the product, assuming the range of the transfer functions is always $f\in[0,1]$. Alternative integration rules such as the mean might have different properties which could influence model identification or performance. For example, a mean integration rule would be differentiable, which allows derivative-based optimization approaches to be used. The particular form of the transfer function could also be explored. We choose a Hill-like function because of its prominence in the systems and synthetic biology community. However, the only mathematical requirement for a transfer function is that it map a non-negative continuous or categorical variable into the range $f\in[0,1]$. Thus, many types of transfer functions are possible."

4. Have the authors investigate the parameter estimation using other combination of measured variables? How "typical" is the parameter estimation result shown in the manuscript?

We thank the reviewer for this insightful comment. Measurement selection (and the design of appropriate experimental perturbations) will be critical when we expand this strategy to larger cell free systems. However, in this study we did not explore alternative measurement sets. We believe the selected measurement set approximated what can be expected in practice, both in terms of network coverage and sampling frequency, based upon published cell free data sets from the Swartz lab, and from current projects in the our lab. In the Swartz study, glucose (starting material), metabolic intermediates such as lactate, acetate, pyruvate, 18 amino acids, energetic cofactors such as ATP and end products were measured (a total of 37 components) at a frequency of approximately 25 min. From ongoing studies in our lab, we expect a typical cell free network to consist of ~200 metabolites. Thus, Swartz and coworkers measured approximately 20% of the network metabolites. In this study, we measured 33% at a sampling frequency of 20 min (opposed to 25 min in the Swartz study). Moreover, our measurement set did not not include highly connected energetic cofactors, which are typically measured. Thus, while only a proof of concept, our measurement set was roughly similar to what we can expect in practice.

Toward the reviewers concerns, we have updated the results section with a reference to the published cell free study of Jewett et al.,

" …This data set is similar to published cell free studies both in terms of network coverage, and sampling frequency."

Jewett reference:
Jewett MC, Calhoun KA, Voloshin A, Wuu JJ, Swartz JR (2008) An integrated cell- free metabolic platform for protein production and synthetic biology. Mol Syst Biol 4: 220.

5. How was noise introduced to the in silico data? From what distribution is the noise? The in silico data shown in Figure 5 did not look to have much noise.

We thank the reviewer for highlighting this important oversight. In the parameter identification studies, we assumed a constant coefficient of variation (CV) of 10% for the synthetic measurements. However, we plotted uncorrupted values for the synthetic data in Fig. 5.

To address this issue, we replaced "… noise corrupted synthetic data…" with "… synthetic data …" in the results section, added an additional line to the caption of Fig 5 to clarify this point:

"… The synthetic data plotted here was unperturbed by noise. However, we assumed a constant coefficient of variation of 10% for the synthetic data during parameter estimation."

And lastly added clarifying text to the materials and methods section regarding the coefficient of variation:

" … We assumed a constant coefficient of variation of 10% for the synthetic data set."

6. Figure 8 is quite busy and confusing since the colors mean different things for different variables. The figure legend does not help in this case. E1, E3 and E6 on the plots should be better written as v1, v2 and v6. Why only show the control variables for these enzymes? To what reference were the metabolic fluxes scaled? The true values should also be plotted (unless all variables are scaled to the true values; but this was not clear).

We thank the reviewer for this helpful comment. Indeed, there is a lot of information in Fig 8, and we agree that this figure can be confusing. Toward this concern, we updated Fig 8 and the caption for Fig 8 to resolve the confusion identified by the reviewer. First, the colors simply mean low (blue) versus high (red) for each of the different components being visualized. We have added a sentence to the caption to better explain the color scale. Second, we changed the E1,E3 and E6 labels in Fig 8 to v1, v3 and v6 as suggested by the reviewer. Lastly, flux j was scaled by subtracting the globally minimum flux and then normalizing this difference by the global (max - min) value. This put the scaled flux values onto a [0,1] scale, where 0 becomes the minimum value, and 1 becomes the maximum value. We have routinely used this scaling in the past in a variety of applications. We have added a description of the scaling to the materials and methods section, along with references to past examples where we have used it.

Fig 8 caption:
"Metabolic flux and control variables as a function of network type and particle index at t = 100 min. The particle error, the control variables governing E_1, E_3 and E_6 activity (v1,v3 and v6) and the scaled metabolic flux were calculated for the positively (top), uniformly (middle) and negatively biased (bottom) particle swarms (N = 100). Blue denotes a low value, while red denotes a high value of the respective quantity being plotted. The particles from each swarm were sorted based upon simulation error (low to high error). **A**: Model performance for the positively biased particle swarm as a function of particle index. **B**: Model performance for the uniformly biased particle swarm as a function of particle index. **C**: Model performance for the negatively biased particle swarm as a function of particle index. Models with significant control mismatch showed distinct control and flux patterns versus those models with the correct or closely related control policies. In particular, models with the correct control policy showed stronger inhibition of E_1 activity, leading to decreased flux from $S \rightarrow P_1$. Conversely, models with significant mismatch had increased E_1 activity, leading to an altered flux distribution. This is especially apparent in the negatively biased particle swarm. "

New references:

Lequieu J, Chakrabarti A, Nayak S, Varner JD (2011) Computational modeling and analysis of insulin induced eukaryotic translation initiation. PLoS Comput Biol 7: e1002263.

Song SO, Chakrabarti A, Varner JD (2010) Ensembles of signal transduction models using pareto optimal ensemble techniques (poets). Biotechnol J 5: 768-80.

7. Figure 7 showed that the parameter angles were similarly distributed for different network structures. What does this imply?

   Please see response to P8.

8. The authors used the term “sloppy ensemble” to describe the existence of an ensemble of structures and parameters describing the data. I do not think that this case was more than an example of parameter estimation with severely incomplete model identifiability (see review Villaverde & Banga, J. Roy. Soc. Interface., 11: 20130505, 2013). The (in)validation of the model ensemble should be done for independent data (i.e. data not used in the parameter estimation). The sensitivity to initial population of network structures also suggested that the model identification was severely ill-posed. This issue, I believe, is a big drawback of the proposed framework. I wish the authors would outline strategies not only for model formulation, but also for model identification, which should include both structure-parameter identification and design of experiments.

   We thank the reviewer for raising this critical point. We used the term sloppy ensemble because we identified parameter vectors that simultaneously gave low simulation error (consistency with measured values) but non-zero parameter angles. Thus, we identified parameter values that were different from their true values, but nonetheless produced reasonable model performance (low error). This suggested that we captured important parameter combinations (stiff combinations), while simultaneously missing other parameter combinations (sloppy combinations). This is similar to the earlier study of Brown and Sethna, which showed that reasonable model predictions were possible, despite sometimes only order of magnitude parameter estimates, if the stiff parameter combinations were well constrained. We have updated the text to better explain this terminology.

   The larger point about an ill-posed identification problem is also very interesting. An exhaustive examination of this question is beyond the scope of the current manuscript. However, we have begun a study using models with structures similar to those presented here to specifically address this issue. We completely agree that we should test the model ensemble using additional data. Toward this, we have suggested in the discussion a strategy to generate the best possible data set to differentiate between competing model structures. When the initial distribution of control structures was unbiased, we observed that the true structure or structures closely related to the true structure were preferentially selected. However, we were often unable to distinguish between the true structure and certain closely related degenerate structures. Degeneracy arose because of the winner takes all integration rules. Other integration rules, such as the mean or product of transfer functions, might eliminate this degeneracy issue. Alternatively, we could also design experimental protocols that maximally distinguished between possible degenerate structures, for example using a strategy similar to Gilles and coworkers. We do have experimental limitations to what is possible, but one could imagine using the model to design pulse chase experiments that would differentiate between competing model structures.

   Toward the reviewers comments, we have updated the discussion section regarding the sloppy ensemble, and added text to the discussion concerning experimental design strategies to distinguish between competing control structures. We added two new references.

First paragraph discussion:
" … Thus, we identified parameter values that were different from their true values, but nonetheless produced reasonable model performance (low error). This suggested that we captured important parameter combinations (stiff combinations), while simultaneously missing other parameter combinations (sloppy combinations). This was similar to the earlier study of Brown and Sethna \citep{Brown:2003aa}, which showed that reasonable model predictions were possible, despite sometimes only order of magnitude parameter estimates, if the stiff parameter combinations were well constrained."

Discussion section (last paragraph):

"… Also, there are many different approaches from the reverse engineering of gene regulatory networks that perhaps could be adopted to this problem, however this remains an open question. For example, one could imagine designing pulse chase experiments which maximally distinguish between competing allosteric models, similar to the earlier work of Kremling et al \citep{Kremling:2004aa}, or iteratively estimate model structures similar to Doyle and coworkers \citep{Gadkar:2005aa}."

New references:
Kremling A, Fischer S, Gadkar K, Doyle FJ, Sauter T, et al. (2004) A benchmark for methods in reverse engineering and model discrimination: problem formulation and solutions. Genome Res 14: 1773-85.

Gadkar KG, Gunawan R, Doyle FJ 3rd (2005) Iterative approach to model identification of biological networks. BMC Bioinformatics 6: 155.

9. In the abstract, the authors stated that "… we could simultaneously estimate kinetic parameters and allosteric connectivity from synthetic data." The results of the model identification however showed otherwise.

We thank the reviewer for highlighting this issue. It is true that if the initial population of particles was heavily enriched with incorrect structures, our PSO approach failed to converge to the correct structure. However, it was also true that we were able to estimate the correct structure and low error parameter vectors when our initial population of particles was unbiased or positively enriched with the proper structure. We have updated the penultimate sentence of the abstract to reflect this distinction:

" …Lastly, we showed, at least for the network architectures considered here, that we could simultaneously estimate kinetic parameters and allosteric connectivity from synthetic data starting from an unbiased collection of possible allosteric structures using particle swarm optimization. However, when starting with an initial population that was heavily enriched with incorrect structures, our particle swarm approach could converge to an incorrect structure. While only an initial proof-of-concept, the framework presented here could be an important first step toward genome-scale cell-free kinetic modeling of the biosynthetic capacity of industrially important organisms."

10. On page 13, remove "was" from "…the P1 inhibition of E_1 was dominate".

We thank the reviewer for this suggestion, and we have corrected the error in the text.

11. On page 15, 4th line from the bottom: Replace the second \sigma_{i,j}>0 with \sigma_{i,j}<0.

    We thank the reviewer for this suggestion, and we have corrected the error in the text.

12. On page 17, please define \Delta. How were the PSO parameter \theta set?

    The following text was added to define Delta after Equations 12 and 13:
    "where Δ_i denotes the perturbation to the vector of parameters k_i for particle *i*."

13. There were missing references for Octave.

    The Octave reference is now provided in the text as (v 3.8.1; www.octave.org).