

Revisions for ‘Bayesian inference and comparison of stochastic transcription elongation models’

We wish to thank the reviewers for their helpful feedback, and the editors of PLOS Computational Biology for considering our paper.

This document is divided into two sections. First, we acknowledge and respond to individual critiques – which have prompted considerable structural changes to the paper. In the second section, we highlight the major structural differences between the original and revised versions.

Relevant changes are also highlighted in the revised manuscript.

We also wish to request two changes in affiliation from the editors. First, the “Department of Computer Science” at the University of Auckland has since been renamed to the “School of Computer Science”. Second, co-author Alexei J. Drummond is now additionally affiliated with the “School of Biological Sciences” at the University of Auckland, alongside the other authors. The new author affiliations are in the revised manuscript.

Issues

1. From both reviewers: the “results” are deferred to the Discussion section. This makes the manuscript difficult to follow.

Response: We agree. The “Results” section has been renamed to “Models” and “Discussion” has been renamed to “Results and Discussion”. The section on approximate Bayesian computation has been moved from the former to the latter. The “Materials and Methods” section has been removed altogether and its material has been dispersed throughout the paper. We hope there is sufficient justification for the editors of PLOS Computational Biology to permit a deviation from the canonical template.

2. The authors should include a summary at the end of the paper.

Response: We have added a “Conclusion” section. This was adapted from what used to be the first paragraph of the discussion.

3. The MCMC does not look like it has converged. \hat{R} and chain lengths should be provided to measure convergence.

Response: The simulations have been run for longer and \hat{R} and chain lengths

have been provided. Additionally, the ESS is now provided for every parameter instead of only providing the range and mean. To assist with interpreting the Bayesian inference, we have plotted the full posterior and prior distributions instead of using a table (new Fig 5). Note that running the simulations for longer did not change the qualitative results of this paper nor did it change the set of models in the 95% credibility set.

4. Chemical master equations should be provided. If these can be solved analytically it may provide deeper insights into the system.

Response: We have provided chemical master equations for single-cycles of the four equilibrium models (see S2 appendix), and incorporated this into our discussion. This information may help readers to better understand the kinetics of the system. Note that the chemical master equations for the full gene sequence involves a number of states that grows with the length of the gene and therefore providing analytical solutions to the chemical master equation becomes impractical. Numerical instabilities would be compounded by the extension of the transcription model to include eg. backtracking, hypertranslocation, nucleotide misincorporation. What we have provided in this paper is a framework for inference on any model of the system.

5. Portions of the manuscript are quite difficult to follow. Figure 1B shows simulations before one has any inkling about the simulation method. The authors spend two paragraphs describing the details of sequence-dependent kinetics and energetics of disruption, which properly belong to the section on model description.

Response: The caption for Figure 1B has been adjusted. The two paragraphs mentioned have been moved into the “Parameterisation of the translocation step” section. The section in “Materials and Methods” concerning nucleic acid thermodynamics has also been moved into this section.

6. Translocation kinetics are difficult to follow. Should state at the outset that the translocation kinetics appeal to transition-state theory. Appendix S3 (four translocation transition state models) provides additional confusion.

Response: We believe that the restructurings described in (5) will help with this issue. Additionally, transition-state theory is now mentioned several times leading into the relevant section. A truncated version of Appendix S3 and its figure have been moved into the “Parameterisation of the transloca-

tion step” section and all mention of the models which are not being used have been removed. We agree that these details were unnecessary and distracting. Finally, we have made various general restructurings of the “Parameterisation of the translocation step” section detailed at the end of this letter. We hope that all of these changes will make the material easier to follow.

7. Figure 4B, which is critical for understanding the models, could use some more information. The parameter being introduced at each step should be specified.

Response: The image has been corrected.

8. Would be useful to summarise results from all models, not just the “best fit” models

Response: The “not-best-fit” models, such as models 1-10 for RNAP and pol II, occur in the posterior distribution with probability zero. This means that given the value of ϵ used in the analysis, the inference engine was unable to find even a single state which has model 1-10.

9. Other minor issues. Page 1, line 12: Change template DNA to DNA. Page 1, line 33: Change practise to practice. Figure 3 legend: Explain why the transition states are irrelevant in a translocation equilibrium model. Figure 2 legend: define β_1 , β_2 , and h here. Page 10, line 302: Use significant figures.

Response: Corrected.

Structural changes

The original (ie. previous) structure is below. Highlighted areas represent new sections / figures or old ones moved to the highlighted place.

1. Introduction

- (a) RNAPs and single-molecule experiments (line 1-26) + Fig 1
- (b) Model parameters and parameter reduction (line 27-43) + Fig 2
- (c) Sequence-dependent kinetics (line 44-62) // Moved this into “Parameterisation of the translocation step”.
- (d) Deterministic vs stochastic models and inference (line 63-83)
- (e) “In this study we...” (line 84-90)

2. Results // Renamed to “Models”

- (a) Notation and state space
- (b) Parameterisation of the NTP binding step // New section. Content moved here from “Model space”.
- (c) Parameterisation of the translocation step
 - i. Thermodynamic models of base pairing energies // New section. (c) from Introduction and “Nucleic acid thermodynamics” from Materials and Methods were moved here
 - ii. Calculation of translocation rates or translocation equilibrium constant // New section. The important concepts regarding translocation kinetic are moved here and the less important details now occur later in the section
 - iii. Energetic bias for the posttranslocated states
 - iv. Polymerase displacement and formation of the transition state
 - v. Energy barrier of translocation // A simpler form of S3 Appendix has been moved here.
- (d) Model space + Fig 4
- (e) Stochastic modelling
- (f) Relation to previous models and stochastic simulations // Sub-section moved here from Discussion

- (g) Model selection with MCMC-ABC + Tables 1 and 2 // Moved into “Results and Discussion”. Further split into two subsections.
3. Discussion // Renamed to “Results and Discussion”
- (a) “In this paper we...” (line 250-263) // Moved to the new section “Conclusion”
 - (b) Relation to previous models and stochastic simulations // Subsection moved into Models section
 - (c) Model selection with MCMC-ABC // New subsection. Adapted from the top half of “Model selection with MCMC-ABC” and “Materials and Methods”
 - (d) Posterior distributions + Tables 1 and 2 + Figs 5 and 6 // New subsection. Adapted from the bottom half of “Model selection with MCMC-ABC”. Table 2 now displays MCMC summaries instead of parameter estimates. Fig 5 is a new figure. Fig 6 was moved here from S3 Fig.
 - (e) Translocation rates differ among RNA polymerases
 - (f) The data does not determine the kinetics of the NTP binding step
 - (g) RNAP has an energetic preference for the posttranslocated state
 - (h) δ_1 may be an important parameter but its physical meaning is unclear
 - (i) Comparing the kinetics of RNA polymerases
 - (j) Bayesian inference of transcription elongation
4. Materials and methods // Section deleted. Subsections are dispersed around paper.
- (a) Nucleic acid thermodynamics // Moved into “Parameterisation of the translocation step”
 - (b) Software and algorithms // Moved into SI and “Model selection with MCMC-ABC”
 - (c) Prior distributions // Moved into “Model selection with MCMC-ABC”

5. Conclusion // Added a conclusion
6. Supporting information
 - (a) Stochastic simulation
 - (b) Chemical master equations // New appendix
 - (c) MCMC-ABC // Moved material here from “Software and algorithms”
 - (d) Approximating the translocation transition state + S1 Fig // Appendix deleted. The 3 transition state models not being used were removed altogether. The 1 model being used (+ S1 Fig) was relocated to “Parameterisation of the translocation step”.
 - (e) Prior distributions + S2 Fig
 - (f) S3 Fig // Moved into “Results and Discussion”