UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY · DAVIS · IRVINE · LOS ANGELES · MERCED · RIVERSIDE · SAN DIEGO · SAN FRANCISCO   SANTA BARBARA · SANTA CRUZ

CALIFORNIA INSTITUTE FOR QUANTITATIVE BIOSCIENCES (QB3)
BERKELEY, CALIFORNIA 94720-3220

John D. Chodera
260J Stanley Hall #3220
Berkeley, California 94720-3220
PHONE: 415.867.7384
EMAIL: jchodera@berkeley.edu
October 26, 2012

Dear *Biophysical Journal* Editors,

We are submitting a revised manuscript for consideration for publication as an **Article** in *Biophysical Journal*:

Bayesian hidden Markov model analysis of single-molecule force spectroscopy: Characterizing kinetics under measurement uncertainty, by John D. Chodera, Phillip J. Elms, Frank Noé, Bettina Keller, Christian M. Kaiser, Aaron Ewall-Wice, Susan Marqusee, Carlos Bustamante, and Nina Singhal Hinrichs.

We apologize for the long period of time that elapsed between the receipt of referee comments and the submission of this revision. The first author, John D. Chodera, was engaged in a protracted faculty search that did not permit proper attention to be devoted to thoroughly addressing the referee comments until now. We request that the referee critiques (included in this letter) and our responses be considered in the evaluation of this manuscript revision, especially considering the time devoted by the referees to providing thoughtful and useful commentary, for which we are highly appreciative. We were also happy to see the enthusiasm the referees share for this work.

Below, we address the referee comments (set in red) with detailed responses (in black), explaining where and how we have adapted the manuscript to meet their concerns. We believe the revised manuscript is much stronger and more broadly accessible, and hope the Editors will find this version suitable for publication in *Biophysical Journal*.

Peer Reviewer #1

This paper applies hidden Markov models to extract thermodynamic and kinetic information from single molecule hopping traces. The authors explain different methods to extract transition rates between an arbitrary number states with Gaussian distributed observables from hopping data.

The problem is certainly very interesting but the reading of the manuscript is difficult and the presentation of the methodology obscure to non-experts. I think this paper should be primarily written in a way that is understandable to an audience of experimentalists working in single molecule biophysics and readers of Biophys. J in general. The core of this manuscript is too specialised (pages 11-18 describe all the steps of the algorithms and are difficult to follow) failing to deliver the main message.

We have reorganized the manuscript to streamline delivery of the main message and make it more accessible to a broad audience, including experimentalists. While we feel strongly that it is essential to present a *complete* description of the algorithms so that other researchers can independently implement them and reproduce our work as closely as possible, we have tried to minimize the impact detailed descriptions have on the main flow for the general reader, so we have relocated such verbose descriptions in the Appendix.

Moreover the experimental part is weak with just one trace of one molecular species extracted from a pool of RNAs that supposedly should only contain the wild type P5ab hairpin. Although I can understand this, it remains unclear whether results are reproducible and the conclusions drawn are tenous: the sequence and the structure of the molecule are unknown and there is no evidence whether the distinct three-state chemical species will appear in another synthesis of P5ab. This paper should start first describing the experimental results and then the theory behind the results in a away that is accessible to experimentalists. Otherwise this paper is suitable for publication in a more specialised journal on optimization algorithms and statistical inference methods.

The p5ab RNA used in this study comes from the same sample used in previous reports [**?**, **?**, **?**, **?**]. While we cannot

rule out the presence of a diversity of molecules resulting from synthesis or storage, we note that the resolution of the instrument used in the current study is superior to that used in previous reports on this RNA system, allowing for the possibility that previous measurements missed the three-state-like behavior we observe in some traces. [JDC: We can show the analysis of two such fibers in the table to demonstrate the results agree within error bars.]

We also note that the main purpose of our inclusion of experimental data was not to draw new conclusions about the behavior of this well-studied RNA hairpin or its potential degradation products or synthesis byproducts, but to illustrate the performance of our Bayesian extension of HMMs on real data containing all of the features of real measurements. In this revision, we therefore include analysis of traces of another system, apomyoglobin at pH 5, a system studied previously [?] on the same instrument. Rate estimation on this system system is especially challenging due to the highly overlapping nature of the force histograms. [TODO: include results on apomyoglobin.]

As suggested, we move up discussion of the experimental measurements earlier in the manuscript to make it more accessible to experimentalists, and add a paragraph summarizing our primary results at the end of the Introduction.

Other comments

Major concerns:

1) What is the potential influence of drift in the method?. In the presence of slight drift the output signal corresponding to different states gets mixed and this may compromise the successfulness of the method. Ideally one could always look for drift-less traces but this may be very difficult in some cases or some experimental setups. The authors should discuss this problem.

Compensating for significant drift solely in analysis is difficult, as changes in the relative optical trap position will change the external potential felt by the molecule, and hence alter the kinetics and thermodynamics. Without imposing a model of how the rate constants and equilibrium state distributions change with the applied trap potential, as well as a physical or empirical model for the dynamics of instrumental drift, it is impossible to compensate for these effects during analysis.

Fortunately, obtaining many drift-free measurement traces of the duration presented here with the generation of miniaturized instrument we employed (where one bead is suctioned onto a pipette and the other bead held in a dual counter-propagating laser trap within a miniaturized, suspended frame) is not particularly challenging—see comments later on this response for further information on how we analyzed drift. Drift is even less problematic for newer generations of instrument that eliminate the pipette altogether, replacing it with another optical trap.

We have added additional discussion to the paper to this effect. [TODO: add discussion about drift to manuscript.] [TODO: Make sure we cite papers on instrument discussing drift (Jeff Moffitt's PNAS paper on dual trap with analysis of S:N ratio, patent, Block papers on dual traps).]

2) How do the finite decorrelation time of the bead in the trap influences the value of the kinetic rates? In the presence of a finite decorrelation time the assumption of uncorrelated (in time) output signal is not anymore valid. Measurements shown in Fig. 3 were subsampled at 1kHz because apparently 1ms is much smaller than the relaxation time of the bead in the trap (bead under tension subjected to the trap and to the tether). The relaxation time (page 2, supplementary information and figure 3 of the Supp. Info) is about 0.5ms (2kHz). Being 1kHz and 2kHz comparable values I do not see how the motion of the bead can be taken as uncorrelated and its effect in the dynamics of the output signal neglected.

The intrinsic timescale or correlation time τ generally indicates that the autocorrelation function of a process has decayed exponentially to $1/e \approx 37\%$ of the original correlation, such that waiting 2τ indicates this process has decayed to $1/e^2 \approx 14\%$ of the original correlation; by 3τ, this is $1/e^3 \approx 5\%$. Because this correlation decays exponentially, it is not the case that 1 kHz and 2 kHz are comparable values. Examination of the power spectrum reveals that the corner frequency of the bead (indicating the timescale for correlation loss by diffusive motion of the bead) is ~2.5 kHz, so $\sim 2.5\tau = 1$ ms $= (1$ kHz$)^{-1}$ results in a decay of this correlation from 100% to $1/e^{2.5} \approx 8\%$, making the bead position within the state effectively uncorrelated at 1 kHz.

See also our response to #3 below.

3) I am concerned about the time decay of the correlation function shown in figure 3 (supplementary). As said in

page 2 (supplementary): "An initial non-exponential or multi-exponential (with unclear number of components) decay is seen in the first 0.5ms to approximately one tenth of the initial value, with slow decay following." Why the initial decay of the bead in the trap should show anarmonicities (leading to multi-exponential behavior)? What is the source of the slow decay observed in the bottom of Fig.3(supplementary)? Surely this cannot be attributed to drift (as this is very small, see my comment 6 below). What is the cause of the apparently "non-ergodic" slow decay? This experimental features of the measurements should be explained.

We apologize for the error in preparing Supplementary Figure 3. We had chosen a trace we had presumed to show single-state behavior from inspection of the histogram (due to the trap position), but further inspection of the temporal trace reveals that this trace in fact shows multi-state kinetic behavior. The long tail of the correlation function is in fact indicative of the infrequent transitions between the kinetically metastable states in this trace, with the initial non- or multi-exponential behavior (which would not be treated by our HMM/BHMM) essentially complete by 0.3 ms, meaning the corresponding intrinsic relaxation times are shorter than this. To clarify the point this figure was intended to communicate, rather than display the correlation function from a different trace that shows single-state behavior, we have added a panel to this figure illustrating the multistate behavior, and amended the caption to point out the short-time decay (not treated by the HMM/BHMM) and long-time decay (explicitly treated by the HMM/BHMM). [TODO: Add additional panel to supplementary figure 3 showing hopping in kinetic trace; plot exponential relaxation fits; amend caption]

4) In page 17 the authors comment on the problem of reshuffling the time data series and the invariance of the full BHMM posterior Eq. 8. How they resolve this question is a bit unclear to me and appears to be a strong limitation of maximum likelihood methods. This problem should be commented much before after Eq. 8 when these methods are first described.

The BHMM posterior is invariant to exchanging labels of the states, not "reshuffling the timeseries"; we have amended the manuscript text to make this issue clearer, as well as moved up discussion of this subtle and somewhat complex issue earlier in the manuscript. This issue does not appear to be a limitation of maximum likelihood methods, but rather simply an issue complicating analysis of MCMC sampling from the BHMM posterior. [JDC: Clarify the state exchange issue and move up discussion.]

5) Page 20, "The P5ab RNA hairpin from Tetrahymena termophilia..... was provided " I would not call the molecule studied P5ab because this leads to confusion. The molecule studied in this work corresponds to a species chemically distinct from p5ab. They should use another name for that species. What is the primary sequence of this molecule? Does it have a contour length compatible with P5ab? Is this new chemical species reproducible in synthesis protocols? What are the secondary structure of this molecule? I have doubts any experimental group whatsoever—maybe neither the authors themselves—can repeat the experiments and test the same molecule again after a new synthesis (??). Why not designing and synthesising an RNA molecule with a reproducible sequence exhibiting a third state with fast kinetic rates? That should not be difficult. Please, remove the term p5ab from the text and figure captions and call it something else.

As the main point of our manuscript is to illustrate the performance of our analysis approach on real experimental data, we have taken two steps to address this issue:

1. We have limited our description of the RNA molecule to an "RNA hairpin" and only describe its origin in the Methods section of the text, so as not to mislead readers that the chemical composition is known to correspond to the p5ab sequence.

2. We have added a section describing our analysis of two-state hopping behavior for another molecule, apomyoglobin at pH 5, whose study by force spectroscopy measurements was recently reported in detail elsewhere [**?**].

6) page 20, "Drift in the instrument was less than 1nm/minute..." Drift is one of the main drawbacks of hopping measurements. 1nm/min looks to me really quite remarkable. One minute is a very long time for a single-trap (rather than dual trap) setup that contains one pipette glued to a glass chamber (that strongly suffers from thermal expansion effects) and a laser beams subject to piezo actuators with creep effects. How many digital-to-analog conversion units do correspond to 1nm in their instrument? I think the authors should show the experimental trace shown in figure 3 filtered at several bandwidth values (10,1,0.1Hz) to show how really big is the drift.

A 32-bit ADC was used where, after calibration, the measured force corresponded to 0.0016697 pN/ADC unit, which is ~600 ADC units/pN. This is now noted in the manuscript.

The average drift in the trap position was determined by analyzing 60-second traces at extreme trap positions where the system was either predominantly folded or unfolded, making few short excursions to the other state. The average force over one second was determined at both the beginning and the end of the trace in a region where there were no obvious transitions. The difference in the force was then divided by the effective spring constant of the system providing the change and therefore the drift in the trap position. Over four such traces, the average drift was 1.1 nm, and the standard deviation 0.7 nm, which correspond to 0.088 pN and 0.056 pN, respectively, given the trap force constant of 0.08 pN. This information is now noted in the manuscript.

We have also added a figure showing filtering (by averaging) at the suggested frequencies to the Supplementary Material. [TODO: Add supplementary figure.]

7) page 21. The large values for k12 and k23 as compared to k13 indicate that a sequential mechanism for hopping is preferred. However the authors suggest that this may not be the only folding pathway. On what basis they can conclude this if they do not know what the molecular structures of the native nor the intermediate are? How can they exclude that indeed the smallness of k13 is just due to numerical uncertainty and that there is just one folding pathway?

The referee is correct that we cannot say for certain what the molecular structures associated with these force states are, so we restrict our language to use the terms "high force," "intermediate force," and "low force." The value of k_{13} is statistically bounded by the 95% confidence interval $[7.2, 11.3]$ s^{-1}, so it is actually quite well determined. However, k_{13} is at least an order of magnitude smaller than k_{12} and k_{23}. If one chooses their criteria for the existence of only one pathway based on a single pathway carrying an order of magnitude more flux than all other pathways, then there would indeed be only one pathway—but this is a matter of how one chooses the criteria for saying so. For whichever criteria are desired, the model provides a statistical way to evaluate this criteria and express its confidence.

Minor concerns:

1) page 5, second line. The acronym MLHMM is not defined.

2) page 5, "...for assessing whether the cost of collecting additional data is outweighed by their benefits." What does this sentence mean? Getting more data is never a problem. Getting good data it is.

3) Eq.1 The variable s_t should be defined.

4) After eq.2 "..that collected data o" "o" should be upper case (rather than lower case)

5) In eq.16 and also below in several places , what does the distance $||..||_2$ mean?

6) a closing parenthesis is missing before eq.22

7) Page 12, "The HMM model.."the word "model" is already contained in the acronym

8) page 15, in subsection "Updating the transition probabilities...." starts with a long, mysterious and convoluted sentence that I could not understand. I suggest to rewrite it in more simple terms.

9) page 17, "We use the (improper) Jeffreys prior..." Why is it improper? Is it because it is non-normalizable? Please, specify.

10) page 19, second paragraph: "As an example synthetic model...high-force folded state". It seems to me this is only true in the passive mode hopping condition (where the force changes in the hopping mode). Please, explain in which mode the synthetic simulation was carried out.

11) page 21, "..the predominant contribution to the observed force....or other sources of measurement noise". I do not quite agree with this statement. Have the authors estimated the expected change in stiffness upon releasing of the folded ssRNA? Did they really expect a marked decrease in the stiffness of the tether upon unfolding?

12) Ref.57 is incomplete.

We thank the referee for their careful attention to detail and willingness to provide these helpful corrections and suggestions where the text could be further clarified. We have modified the manuscript accordingly.

Peer Reviewer #2

This manuscript addresses the highly relevant issue of how to extract accurate kinetics from single force spectroscopy experiments, as well as on how to obtain reliable error bounds, using a Bayesian approach. The method is developed in detail, and then applied to synthetic data as well as to optical tweezer data on an RNA hairpin.

The paper is very well written, and the mathematical treatment is very clear and convincing. I'm convinced the proposed approach is a significant advance over the established methods used to interpret force probe data, and the authors are to be applauded for this. Personally, I'd very much like to see this method published in BJ, as in my view this and similar types of methodological advancement forms an essential part of Biophysics. However, current Editorial policy unfortunately requires each single paper to immediately present significant new biological insight, which is very marginal in the present manuscript. Due to this unfortunate restriction, BJ will probably not be able to enjoy the well above average number of citations this paper will undoubtedly get in the long run.

We thank the reviewer for their high praise, and hope the Editor is persuaded our paper is sufficiently exceptional to warrant publication in *Biophysical Journal*.

Peer Reviewer #3

This paper is worthy of publication after some flaws are corrected. These flaws are not all minor (see full review), but I do not expect them to pose any insuperable problems (hence my belief that the paper can already be judged worthy of publication).

Chodera et al present a new approach to analyzing single-molecule force spectroscopy data, using a Bayesian extension to the classic hidden Markov models that have been used successfully by others. What's new here is the use of Bayesian inference, which provides good estimates of the uncertainties and allows the usefulness of extra data collection to be assessed. The authors also point to a number of useful extensions of the scheme at the end of the paper, such as ways to determine the number of states automatically. The authors describe the conceptual and mathematical bases for the model, demonstrate it on simulated data, and finally test it with experimental data.

The manuscript is generally well-written (though see points below). It presents a worthwhile advance in analytical methods which is likely to find application to a range of single-molecule experiments of interest to the biophysics community. It is therefore a good fit for the Biophysical Journal.

We thank the referee for their support of the publication of this work in *Biophysical Journal*.

However, there are a number of points which should be addressed by the authors before the manuscript is ready for publication.

Major issues:

1. The authors motivate the work in the context specifically of force spectroscopy measurements, where the trajectory of the force is plotted as a function of time for a system under equilibrium. The method seems to be much more general than this, however, and presumably would be applicable to FRET traces, ion channel recordings, and other such single-molecule measurements. Is there some reason that the authors have limited the discussion of the applicability of the technique to force spectroscopy? If so, the reasons should be discussed in the manuscript. Otherwise, the appeal and utility of the manuscript would be greatly enhanced by a discussion (even if brief!) of how the method can be applied to other types of measurements (and any pitfalls that might need to be overcome with other types of data).

The referee is correct in that our Bayesian HMM extension can, in principle, be applied to other single-molecule experiments such as FRET traces. There is, however, a key distinguishing feature between force spectroscopy and these other measurement methods: In force spectroscopy, the observables (force or extension) are continuous numbers on the real number line, while in the case of FRET traces, instantaneous photon arrival times and colors (often along with other properties) are recorded. This necessitates binning the data in order to produce a continuous num-

ber, a procedure that does not retain all of the information encoded in the photon arrival times. Indeed, this binning procedure has been shown to reduce the amount of information that can be extracted from the data [JDC: insert reference from Frank Noé here], which is why we feel we should not recommend this binning procedure. Instead, we focus on force spectroscopy, where the measurements are inherently performed on a continuous observable. We now note this in the manuscript. [JDC: Note this. Problem also occurs for ion channel recording.]

2. Whenever a new method is demonstrated, it is best to show that it "works" by testing it on a dataset for which the results are known from previously-established methods, and then to demonstrate it for data that can't be analyzed well by previous methods. The authors have effectively done the first (with the simulated data, though this isn't as stringent a test as using real experimental data, see point 4), but they haven't quite done the second–they still need to show that their method works on datasets where others don't. They could show this easily using the dataset in Figure 3, for example comparing to the results obtained when the states are separated into different regions of force.

We have added a much more difficult test of the BHMM method for a two-state system with strongly overlapping states, and show how the standard analysis technique of dividing the states by a threshold and estimating the rate constant from the number of threshold crossings fails to recover the true rate, while the BHMM method recovers the correct rate estimate and statistical uncertainty in this estimate. To ensure that we can rigorously test the accuracy of the method, we again use synthetic data for this test. TODO: new synthetic data test and comparison with thresholding

3. More specifically, since this paper is about an extension of classic hidden Markov modeling, the authors should provide at some point in the manuscript or supporting information a comparison of the results from the new approach to the results from a standard HMM model (both the values and uncertainties). Some discussion about comparative computational cost would also be helpful.

We have added a discussion of and comparison to the only other practical approach to estimating uncertainties from standard hidden Markov models, which is bootstrapping based on the generation of synthetic data from the MLHMM. [TODO: comparison with HMM bootstrapping.]

4. In the simulated data, the authors seem to have included Brownian noise but not drift. Since drift is often a problem with experimental data (not all of which is as nice as the data shown in Fig 3!), the authors should show how robust the method is when dealing with drift that is not insignificant (say, half the spacing between the two higher-force peaks in Fig. 2).

The method we describe will obviously fail in the presence of drift as large as suggested, but it is fortunately straightforward to collect data with sufficiently small drift on current-generation instruments that this is not a significant limitation [?]. As an aid to the readers, we have added a section in the Supplementary Material illustrating the behavior on a simple two-state system as a function of drift, so that the reader may more easily understand the pathologies that develop when large amounts of drift are present. TODO: drift discussion in supplemental info

[JDC: Point out that modern systems do not have a problem with drift.]

5. Another issue with the simulated data and the comparison to the threshold method (Supplementary Material) is that the lifetimes in the simulated data appear to be long enough compared to the filtering window that was used (the shortest lifetime, 10 ms, is many times longer than the 1 ms interval time) that the simulated data could very easily have been filtered enough to reduce the overlap between the states to the point that the threshold method works equally well. Hence this is not a particularly impressive test of the method, nor is it a demonstration that the threshold method doesn't work. The authors should be demonstrating the method for a case in which it would simply not be possible to apply the threshold method.

We have chosen the additional "demanding" test described above (in response to point #2) to address this request as well.

6. The authors mention that the experimental data they use to demonstrate the method, in Figure 3, come from RNA molecules that do not display the "usual" two-state behavior previously reported, but instead a different three-state behavior. This raises a major concern because it seems likely that the folding behavior is no longer the intrinsic behavior of the RNA, but may instead reflect damage to the RNA (for example, during storage as mentioned by the

As discussed in the response to comments from the other referees, we have made it clear that the main point of this paper is to illustrate the performance and utility of the proposed procedure on real data, and so the choice of systems studied is secondary. We have clarified our description of this system as well as added another system (apomyoglobin at pH 5) recently described in another paper [**?**].

The issue of determining how many states are best supported by the data is a complex one, and outside the scope of the current manuscript, which demonstrates a procedure for inferring the rates, state descriptions, and uncertainties given that the number of states has been decided upon by the experimenter. We discuss extensions of this approach to decide the appropriate number of states, but such extensions are sufficiently complex that it would double the manuscript length, as well as the amount of time required to develop the methodology and code. The issue of how best to decide the appropriate number of states supported by the data is a matter of ongoing research in the field, and we have done our best to describe this in our Discussion.

We have amended the discussion in the suggested manner. [TODO]

We hope the reorganized manuscript (where much algorithmic detail has been moved to the Appendix and illustrative examples have been moved forward) makes the revised manuscript much more accessible to a general biophysics audience.

The first approximation previously described was indeed motivated by the threshold approach, and as suggested

by the referee, can be used to provide a point of comparison between the initial (threshold) model, the maximum likelihood HMM, and the Bayesian HMM. [JDC: Mention if we decide to do this for synthetic data in the SI.]

To increase the robustness of our approach in handling pathological situations noted after initial submission, we now use a new initialization approach in which the likelihood maximization approach is initialized multiple times from randomized initial models, and the model with the highest log likelihood selected for the maximum likelihood HMM. This new procedure is now described in the manuscript.

[JDC: Make changes to manuscript and code.]

14. Line 312: the authors should explain what is improved over the initial approximation, and why.

15. Line 326: the authors should explain what "Algorithm 1" does, and why it is used.

We have taken these suggestions and made appropriate corrections.

16. Figures 2 and 3 are very difficult to read. The authors should show a zoom into a small region of the graphs, so that the readers may see the transitions more clearly. The figures would also be easier to parse if the individual data points were connected by lines, so that readers could see the data as a trajectory that evolves in time rather than scattered points. This would also make the transitions between states clearer.

We strongly disagree regarding the connection with lines, as this can deceive the eye into perceiving features that do not exist in the data, which is more properly represented by ordered points.

We do, however, now provide zoomed regions of these plots as Supplementary figures. [TODO: supplementary figure]

17. Show Figure 3 in the Supplementary Material on a log scale (the standard way to do it for correlation functions, since the exponential decays can be seen more readily).

We now show Supplementary Figure 3 on both normal and logarithmic scales. [TODO: supplementary figure]

Kind regards,

John D. Chodera,
corresponding author