Experiment by Clark Fritsch, with coauthors Arpan Bhattacharya, Martin Ng, Hong Li, Barry S. Cooperman, Yale E. Goldman

Single-molecule evidence for a new proofreading step in ribosome

Phil Nelson, Univ Pennsylvania

For these slides see:
www.physics.upenn.edu/~pcn

Plan

- Why
- Experiment and puzzle
- Revised proposal for kinetic cycle
- On fitting
- First-passage times: second binding
- Results

Background: It matters

- The ribosome has various "proofreading" steps. They've been studied closely in bacterial ribosome – ***not so much yet in eukaryotes.***
- There are ~7000 genetically transmitted disorders, ~ 11% of which are nonsense mutations like cystic fibrosis (CF), Duchenne muscular dystrophy (DMD), etc. which specifically involve transition of a valid codon to a "stop" codon.
- Symptoms in patients can be alleviated even with just small amount of full length protein.
- Drugs such as Ataluren hold promise for helping ribosome to chug through this particular "stop." How do they work?

I won't answer, but it would be good to know as much as possible about the working cycle of the eukaryotic ribosome.

Shalev, M. and Baasov, T. (2014) Med Chem. Commun, 5(8):1092-1105. Loudon, J.A. (2013) J Bioanal Biomed, 5:079-096. Nadeem Siddiqui, and Nahum Sonenberg PNAS 2016;113:12353-12355

Actual message of this talk

- Sometimes your model's prediction is a probability distribution (really always);
  and
- Single-molecule biophysical techniques give you individual data points for individual molecular transactions;
  but
- Many of us grew up binning data, then least-squares fitting it, which destroys some of its information content, distorts relative importance of different parts of the data, etc.
  so
- The fact that that that's *often unnecessary* is potentially interesting, even beyond the scope of today's application.

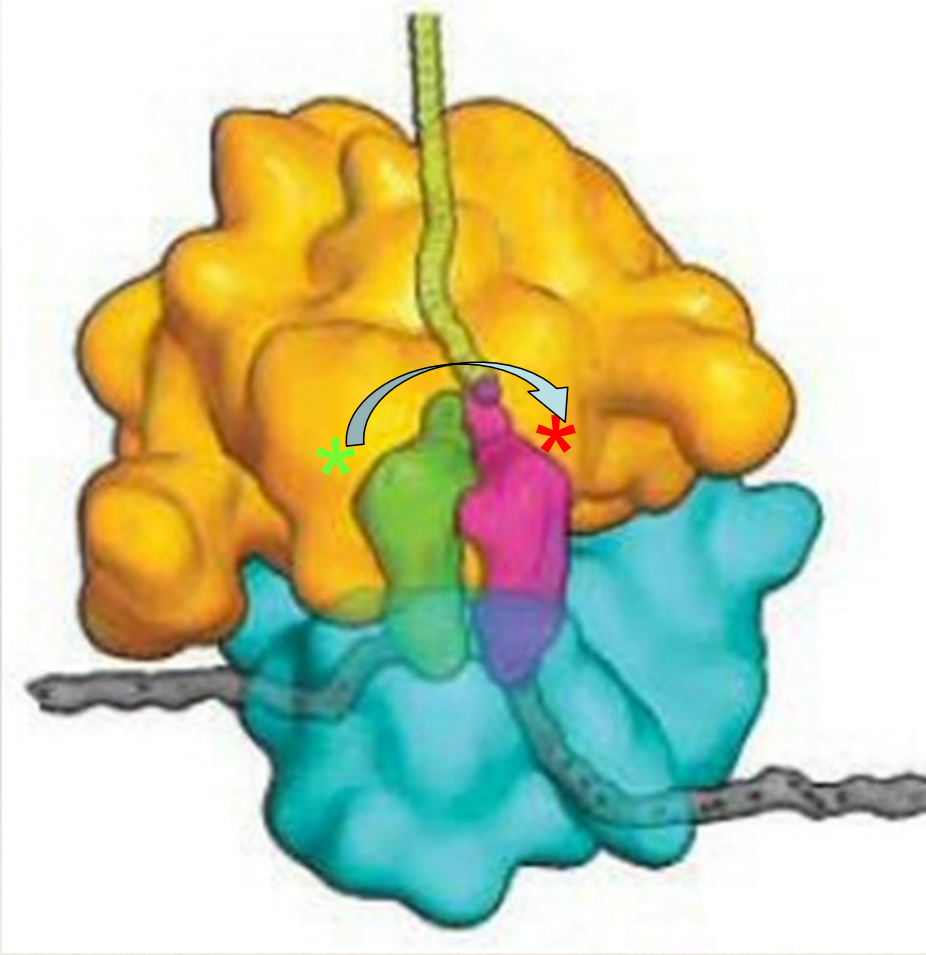
PN, *Physical models of living systems 2nd ed.* (2022).

Plan

- Why
- Experiment and puzzle
- Revised proposal for kinetic cycle
- On fitting
- First-passage times: second binding
- Results

Experiment

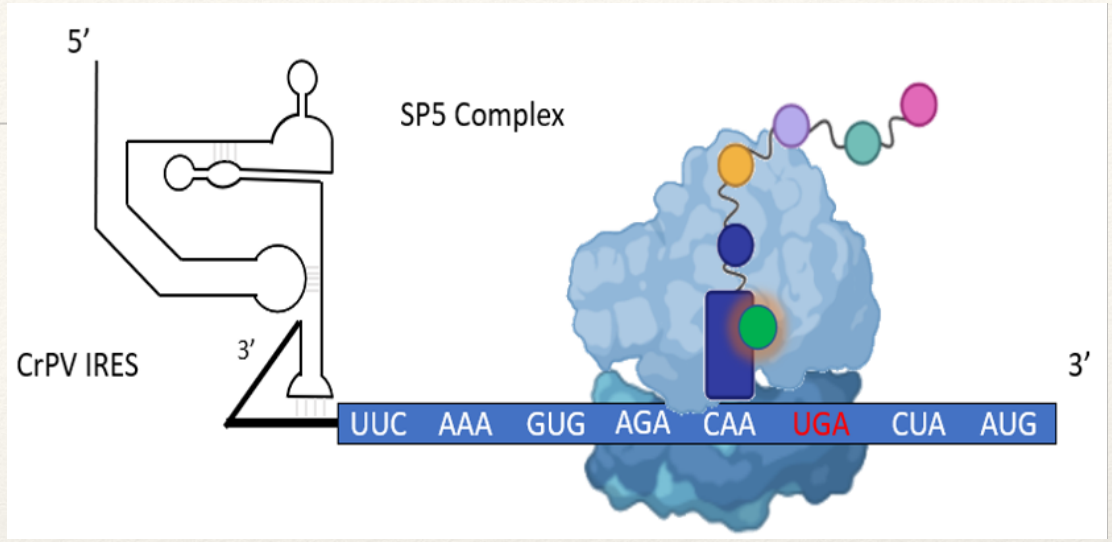
Single-molecule Fluorescence Resonance Energy Transfer (smFRET) tells exactly when two specifically labeled molecules are spatially close (high transfer) or not (low transfer). Hundreds, even thousands of molecules can be simultaneously monitored yielding individual time courses.

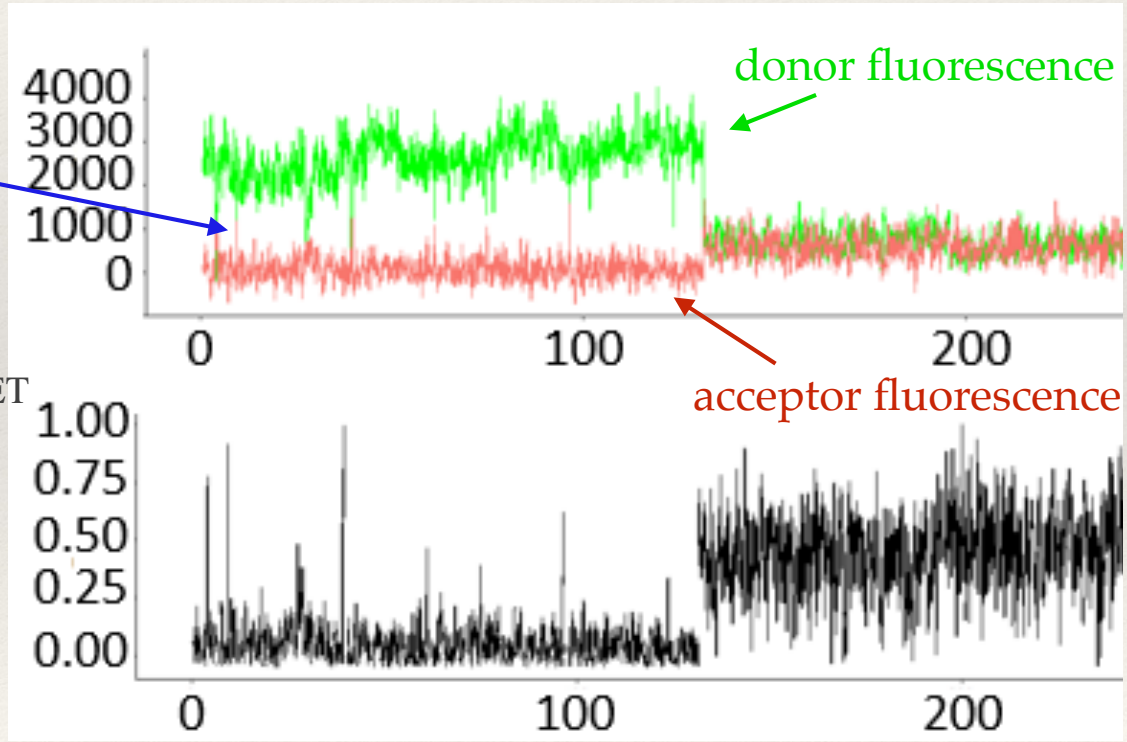


Schematic showing ribosome assembled on an mRNA with a UGA (stop) codon positioned in the A-site of the ribosome. To visualize every binding event, my colleagues made a FRET pair consisting of ternary complex in solution with a donor fluorophore and already-incorporated tRNA in the ribosome with an acceptor fluorophore in the P-site.

Experiment

tRNA is supplied in solution in the form of "ternary complex," or "TC." In experiments I'll discuss today, only one, slightly-wrong (*"noncognate"*), TC is supplied. It samples the A-site of ribosome, binding transiently until eventually it is (wrongly) bound stably. *FRET lets us see individual binding and unbinding events with high time resolution.*

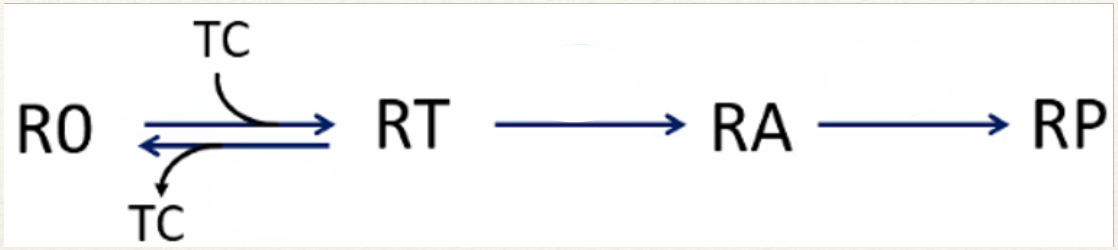




Representative single-molecule trace collected to study eukaryotic tRNA selection on ribosomes programmed on a near-cognate mRNA.

Experiment by Clark Fritsch, Arpan Bhattacharya, Martin Ng, Hong Li, Barry S. Cooperman, Yale E. Goldman.
[Changepoint detection was performed with standard algorithm and is not our subject today.]

Uh-oh



In the simplest model, of course initial binding should be faster if ternary complex (TC) is more abundant. But every binding event is predicted to be independent of every other one, and in particular:

- The distribution of waiting times to bind near-cognate TC should be the same for every attempt.
- The distribution of the number of attempts before stable binding should be independent of TC concentration.

Both of those predictions were found to be false.

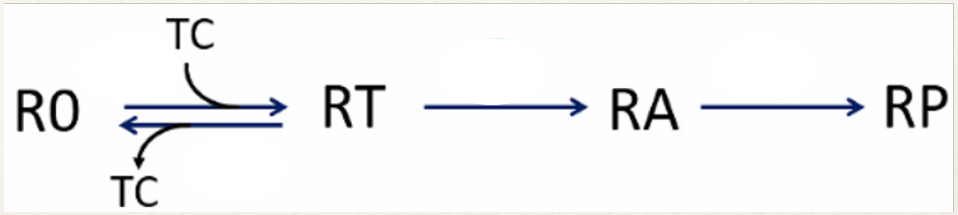
- The distribution of waiting times for near-cognate TC to bind the first time (single exponential) was *qualitatively different from subsequent times* (double exponential).
- The mean number of attempts before stable binding of near-cognate TC was an *increasing function of ternary complex concentration.*

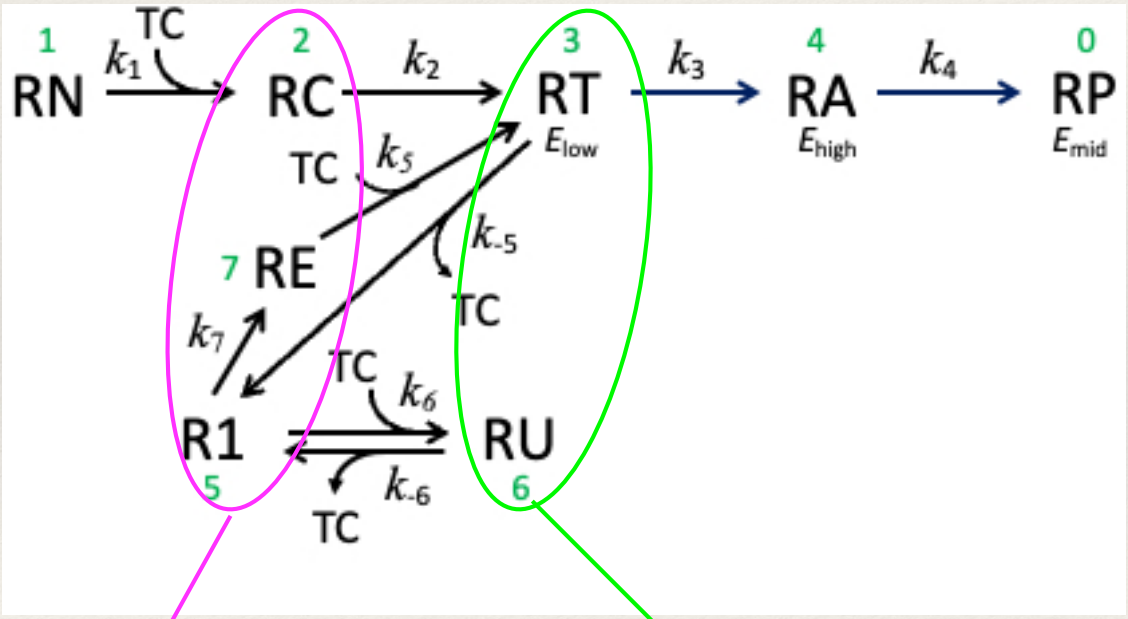
Plan

- Why
- Experiment and puzzle
- Revised proposal for kinetic cycle
- On fitting
- First-passage times: second binding
- Results

Revised proposal for kinetic cycle

Instead of:





zero-FRET (unbound) states

higher-FRET (bound) states

Hypothesize a new side-branch with a dead-end, as the main route for the tentatively bound ternary complex to be rejected from RT.

I won't attempt to argue for this model; I will try to work out its experimental signatures.

Plan

- Why
- Experiment and puzzle
- Revised proposal for kinetic cycle
- On fitting
- First-passage times: second binding
- Results

Max-likelihood fitting

The power of single-molecule methods is that instead of ensemble-averaged data, we have the *actual duration of every* binding event, thousands of them (and similarly unbinding events).

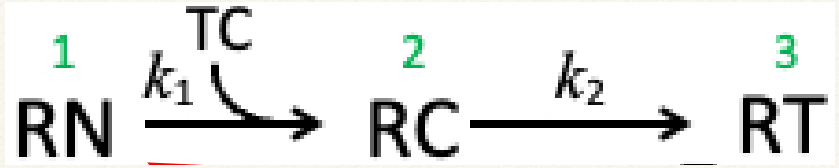
If we can get a model to predict the probability density function of those durations in terms of a few parameters, then we can compute the likelihood of an experimental dataset in terms of those parameters, that is, the probability that the data we *did* observe *would have been* observed in a world with certain values of the parameters.

Then we maximize over parameters, holding the data fixed, to get the values best supported by the data.

It's easy in the familiar case of the release step, a simple exponential. Let's instead look at the more challenging parts of the problem.

PN, *Physical models of living systems 2nd ed.* (2022).

Initial binding as MM



In the model, first binding is a Michaelis–Menten process. Luckily, smart people have already worked out the PDF of completion times: Let $\kappa_1 = k_1[TC]$ and

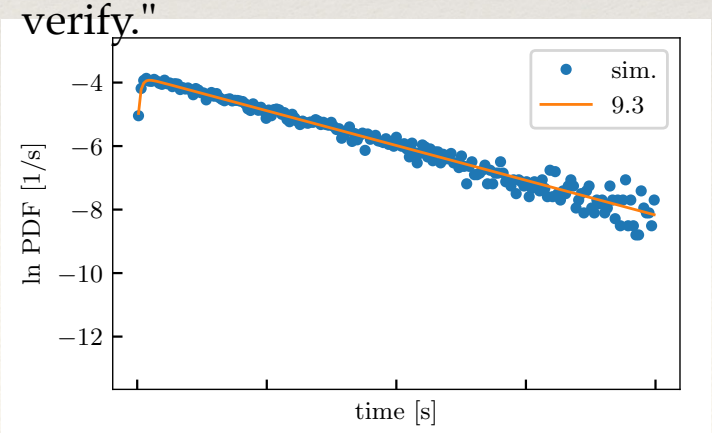
$$B' = (\kappa_1 + k_{-1} + k_2)/2, \;\; A = \sqrt{(B')^2 - \kappa_1 k_2}$$

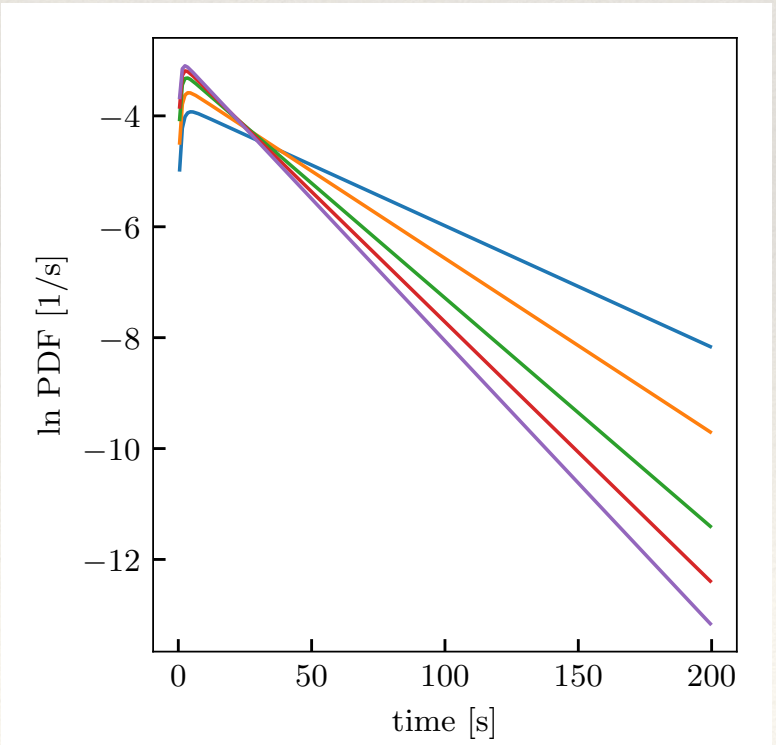
$$\wp(t_1) = \frac{\kappa_1 k_2}{2A} \mathrm{e}^{(A-B')t_1} \left(1 - \mathrm{e}^{-2At_1}\right).$$

Kou et al, J. Phys. Chem. B 2005, 109, 19068--19081.

This distribution can be used to define a likelihood function that determines k_1 and k_2 from t_1 data. We measured several sets of t_1 values, each with a different, but known, [TC].
[*First unbinding* is easier—no concentration dependence.]

Right: This PDF is shown for various [TC] and illustrative *k* values.
Below: It's easy to confirm the result by simulation. "Trust but verify."

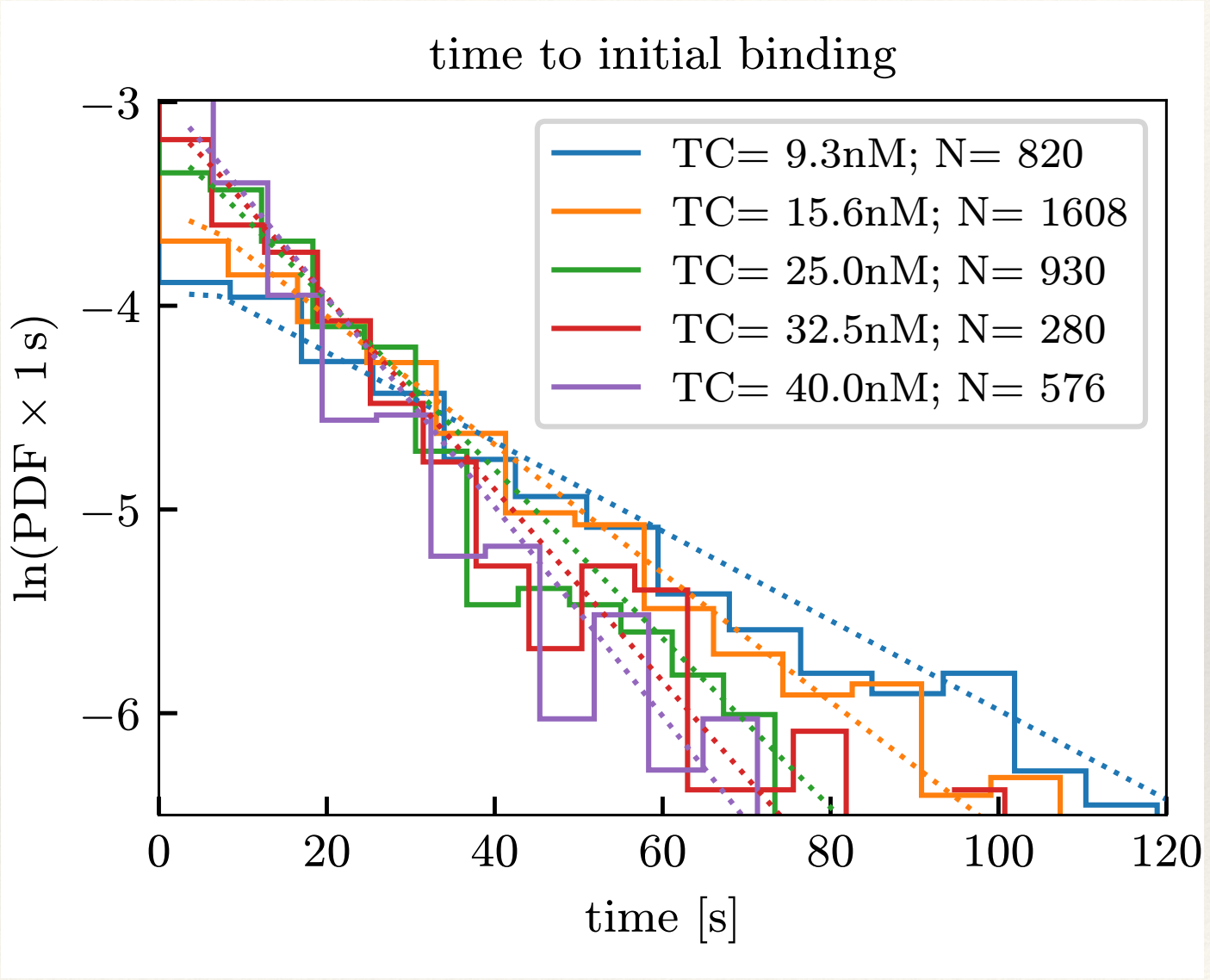




MM can explain first-binding data

So great – we got our model to divulge its PDF. Can the model actually fit real experimental data? It's a tall order – lots of data, just a few fit parameters to get a global fit. – highly overdetermined, which means highly falsifiable.

Initial binding looks pretty good: just two rate constants fit a lot of data.

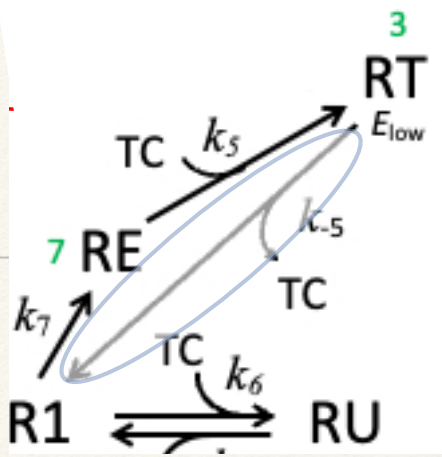


Clark Fritsch, et al., in preparation 2024

Plan

- Why
- Experiment and puzzle
- Revised proposal for kinetic cycle
- On fitting
- First-passage times: second binding
- Results

Subsequent binding as first-passage



Every unbinding brings us to state R1. Because RT and RU are both high-FRET states, we want the distribution of the first-passage time to *either one*. This one I had to work out for myself.

This time the formulas are too long to decently display, but it comes down to convolving two steps for the upper pathway, then finding probability per unit time for first arrival at either R1 or RU, given that the event is "sampling," that is, known to not be the final binding.

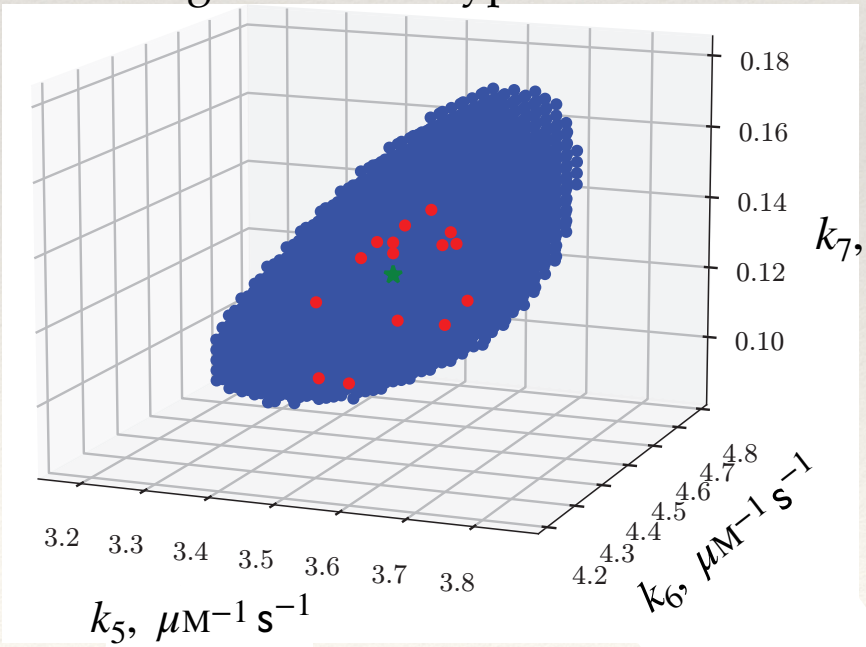
We are using *all* the data in the likelihood function. If we wish, we can then look at reduced statistics to get a human-viewable look at some aspects of our fit.

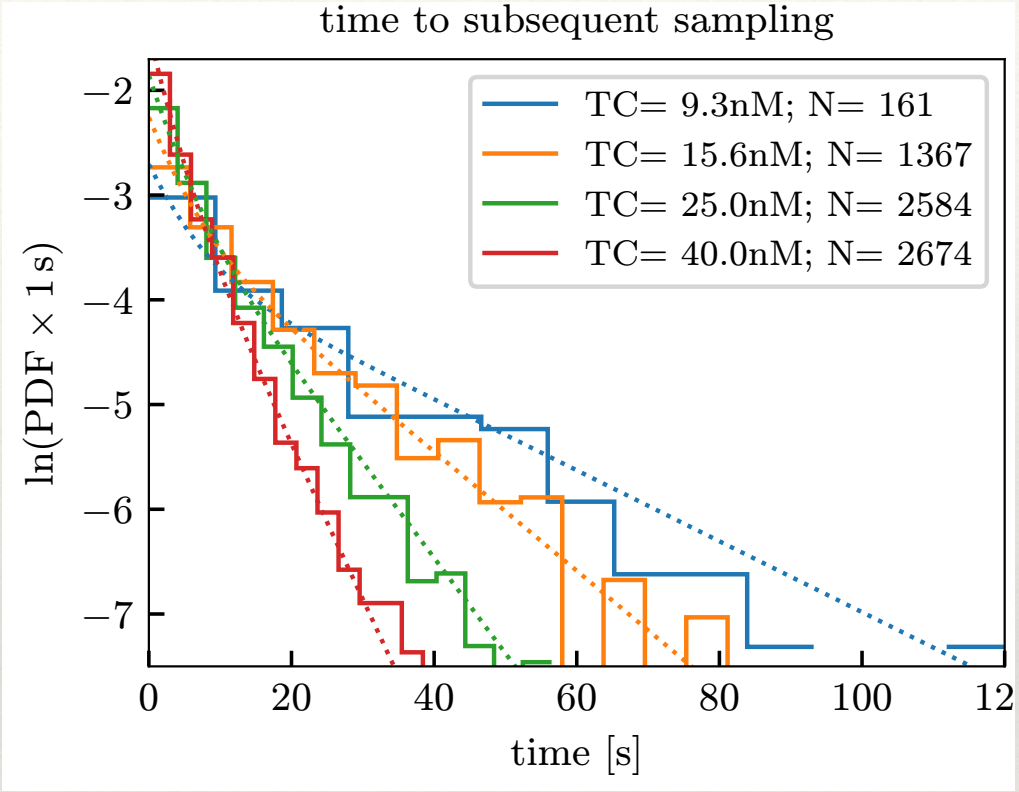
Model can explain subsequent binding data

It's good to have a lot of data, so that we can see deep into the telltale tails of the distributions – the transition from one exponential to the other.

Beyond answering our first paradox, this highly overdetermined fit is quantitatively pretty good, and determines more rate constants (green star below).

Moreover, bootstrap replicates of the experimental data (red) define a cloud of credible rate values that *excludes infinity* and hence argues for the hypothesized new state:

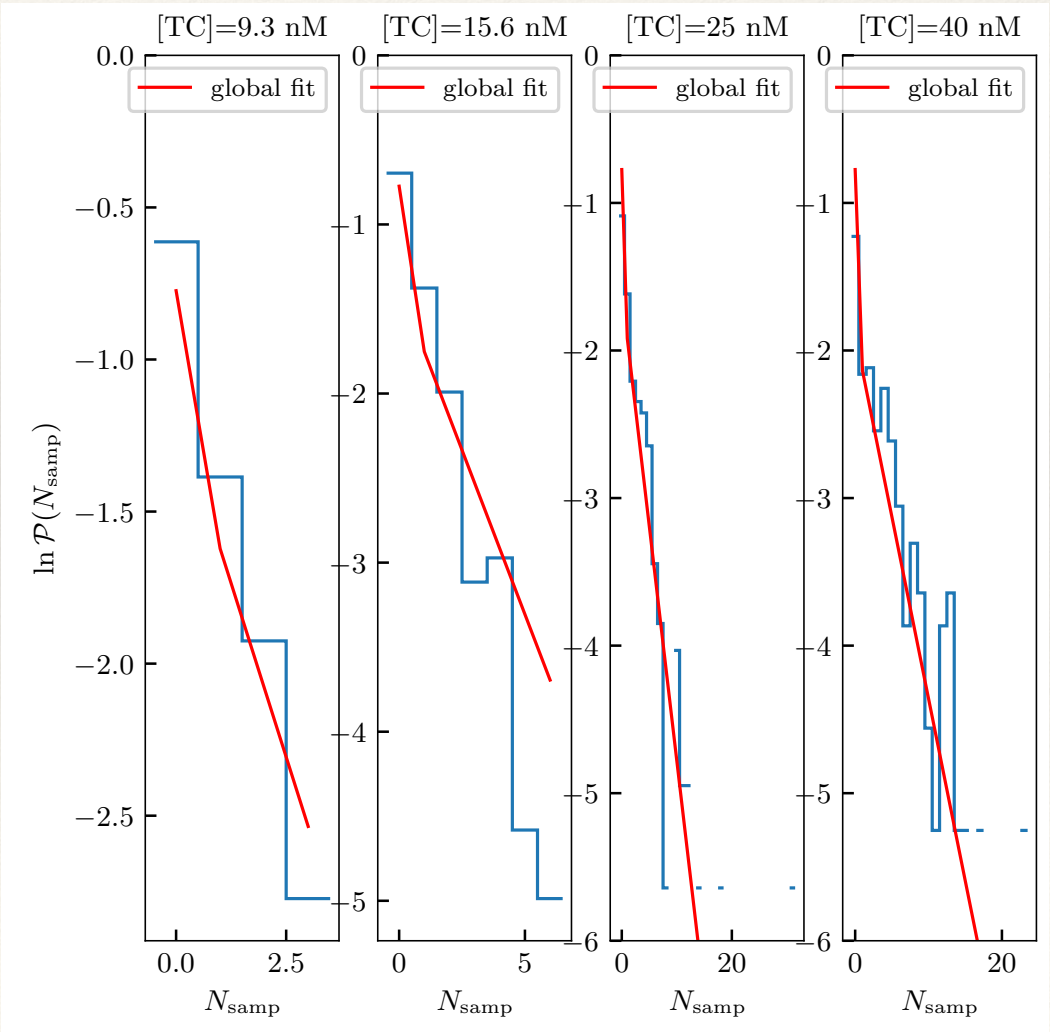




Clark Fritsch, et al., in preparation 2024

An acid test

But once we've found our best version of the model, can it *also* explain *other, different* phenomena that it *wasn't* trained on? We asked it to predict in detail the probability distribution for the number of attempts before stable binding. With no additional adjustment we got:



Beyond answering our second paradox, this highly overdetermined fit is quantitatively pretty good.

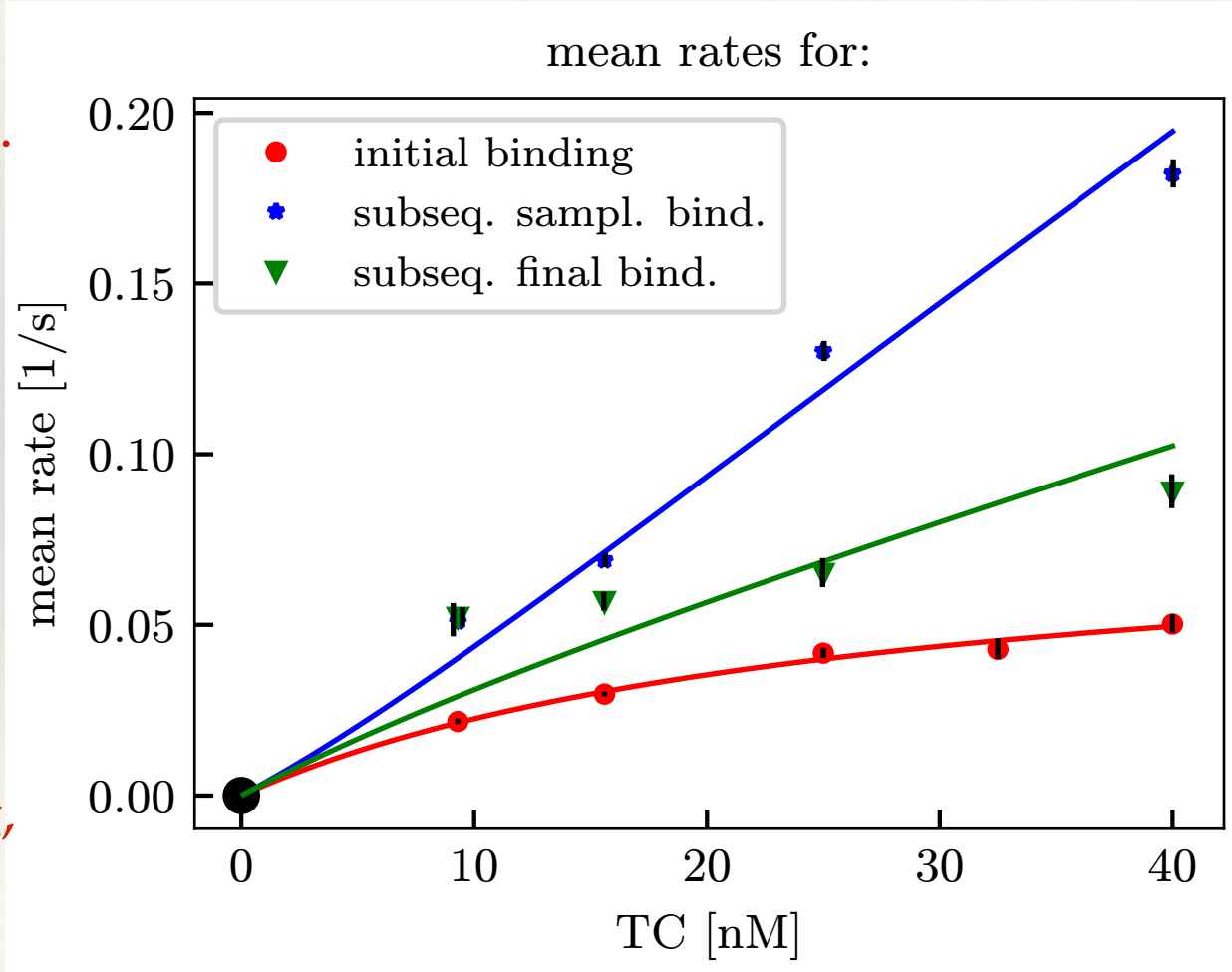
Clark Fritsch, et al., in preparation 2024

Summary

Let's return to the the qualitative observed, surprising, phenomena that motivated the model:

- The distribution of mean waiting times for near-cognate TC to bind the first time was different from subsequent times. **Looks good ---->**
- Not only the first moment (mean rate) but also the entire PDF of waiting times looked good.
- A completely different distribution (of attempt numbers) was successfully predicted with no further fitting, including its puzzling trend.

Clark Fritsch, et al., in preparation (2024)



Thanks

These slides will appear at
www.physics.upenn.edu/~pcn

- The textbook model for ribosome cycle cannot account for new experimental data on eukaryotic ribosomes.
- A minimal extension of that model can.
- The proposed new steps may be a proofreading scheme.
- No kinetic model is ever complete, of course.

University of Pennsylvania

NSF CMMI--1548571

CEMB Center for Engineering MechanoBiology

About likelihood maximization and more: new 2023 edition