Brief summary modeling

Misha Klein   Behrouz Eslami   Martin Depken

THE FREE-ENERGY LANDSCAPE

We build a physical model of Cas-gRNA binding to its DNA substrate. The guide-loaded protein can either be unbound - being in solution - bound to the PAM or having formed an R-loop (intermediate) of a length between 1 and the guide-length (20nt for Cas9) (see figure 1). Furthermore we envision a 'zipper-like' process in which the R-loop grows/shrinks with single nulceotide increments/decrements. The model is therefore fully described by the set of 21 forward rates and backward rates. These rates could in principle depend on both the length of the hybrid formed, termed 'position dependence', as well as sequence content. However, for now we attempt to find overarching targeting behaviour, which should be sequence independence. With this in mind, we assume there are just 3 distinct forward rates: a binding rate from solution to PAM, an internal forward rate while extending the R-loop and a rate from the PAM bound state to initiate the R-loop. The latter is to a good part insired by the bulk data shown by your group implying that initiation the R-loop is the rate limmiting process for cleavage.

To obtain the backward rates, we chose to model the free-energy of every intermediate state and use the detailed balance condition to relate those - via the forward rates just mentioned - to the desired backward rates.

$$k_{\mathrm{b}}(i) = e^{+\epsilon(i)} \times k_{\mathrm{f}}(i-i) \tag{1}$$

with:

$$\epsilon(i) = \begin{cases} \epsilon_{PAM} & \text{at PAM} \\ -\epsilon_C(i) & \text{for a match} \\ \epsilon_I(i) - \epsilon_C(i) & \text{for a mismatch} \end{cases} \tag{2}$$

Given a set of energy differences that make up the free energy landscape (see figure 2) and the three remaining forward rates, we can set up the appropriate evolution equation for the set of populations in each of the states of our model:

$$\frac{\mathrm{d}\vec{P}(t)}{\mathrm{d}t} = M\vec{P}(t) \tag{3}$$

Solving this set of Master equations essentially gives us access to any observable quantity of the system.
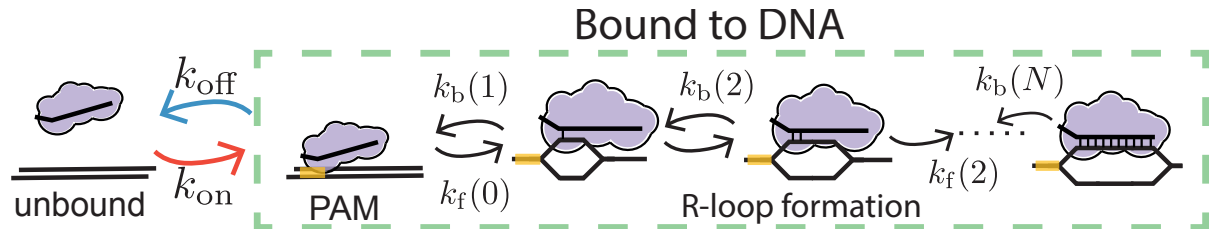

FIG. 1: *Kinetic model of dCas9 (un)binding dynamics. On the microscopic level, we treat every configuration with a single basepair added/removed to the R-loop as a (meta-)statble intermediate. On the more coarse-grained level, all states in which any kind of R-loop is formed, or the protein is attached at the PAM site, are considered to contribute to the population of bound molecules ($P\mathrm{bound}(t) = \sum_{n='PAM'}^{N} P_i(t)$).*
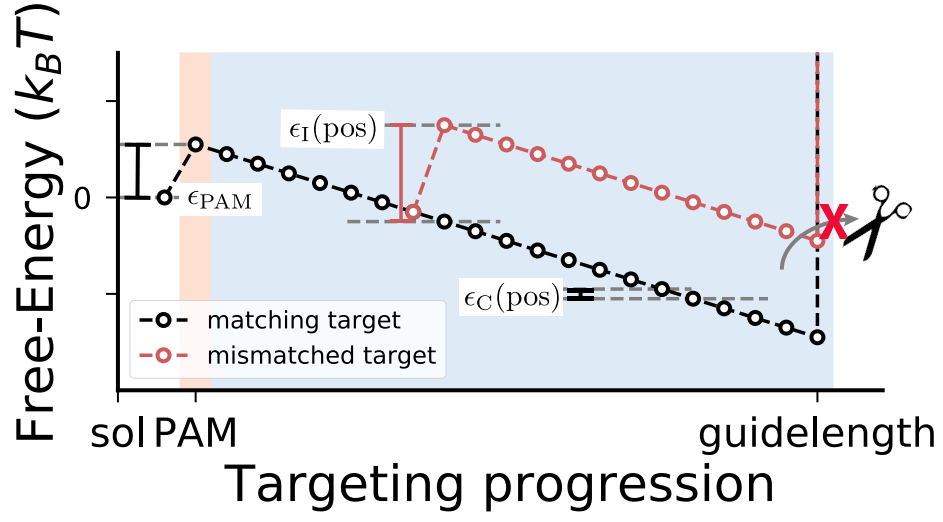
FIG. 2: *microscopic free-energy landscape for the active enzyme. Parameters are 20 gains/penalties for matches* ϵ_C, *1 energy of the PAM bound state* ϵ_{PAM}, *20 mismatches penalties* ϵ_I *and 3 distinct-valued forward rates (binding rate from solution to PAM, the rate from PAM to enter the R-loop and the rate of adding a basepair to the R-loop).*
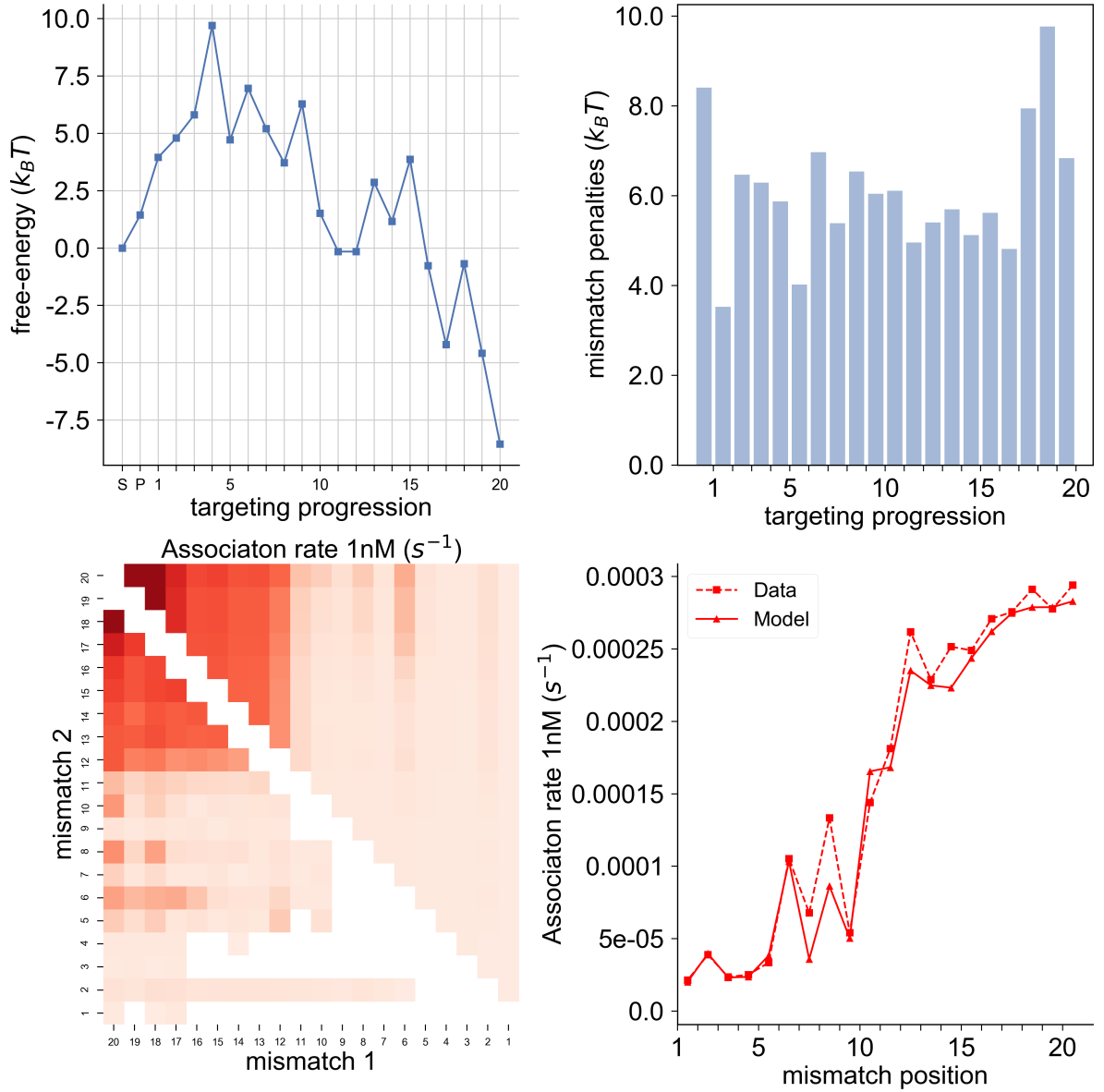
FIG. 3: *Costum build simulated annealing algorithm to fit our model to experimental data. Shown are: (top left) On-target free-energy landscape, (top right) mismatch penalties, (bottom right) comparison to single missmatch data and (bottom left) comparisson to all double missmatch data.*
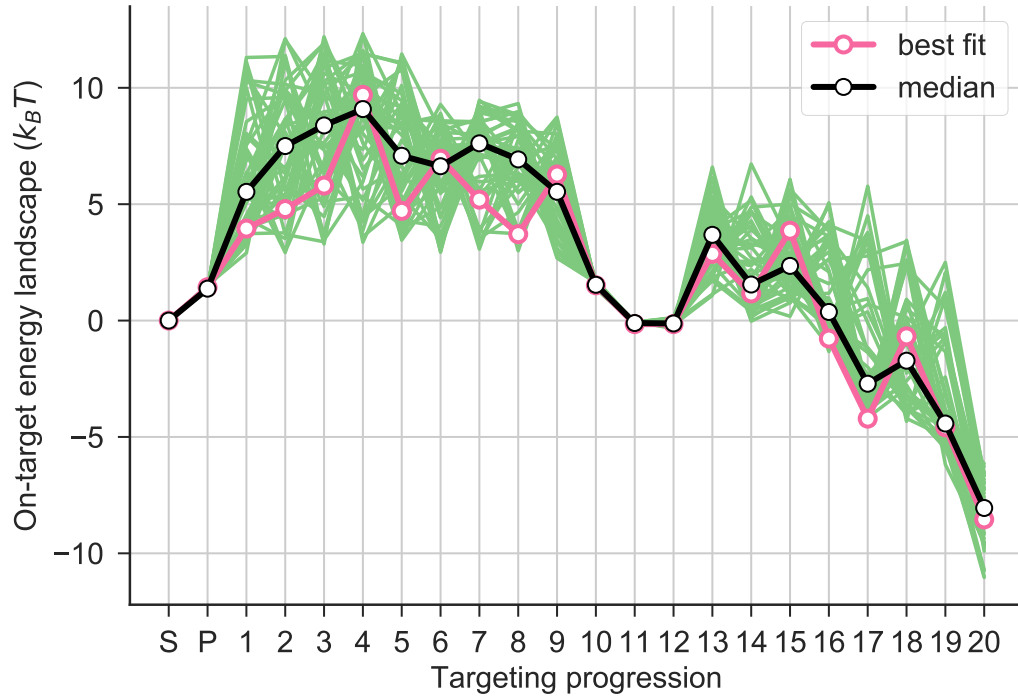
FIG. 4: *On-target free-energy landscape after 66 repititions of the SA-algorithm. By overlaying the results from multiple fits - that are all equally good in performance (< 5% difference in predicted associated rates)- we see what features are very strongly determined by the data and which are not so much.*
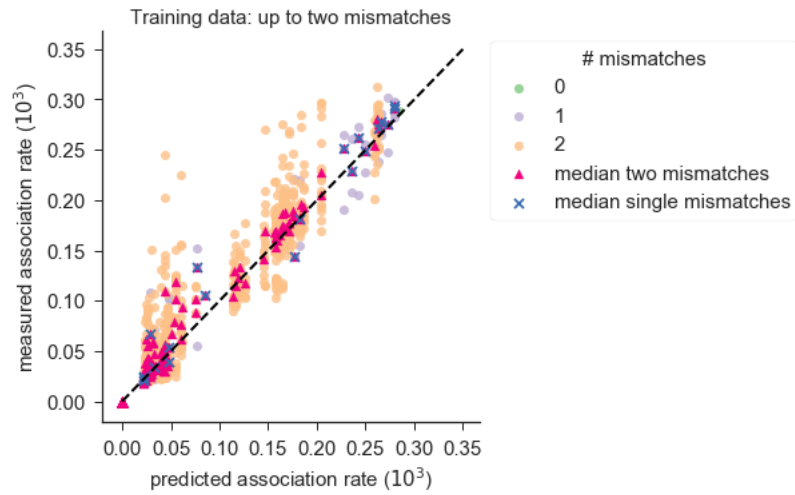

FIG. 5: *Correlation between training data and model prediction: 0.94367427*

TESTING USING DATA FINKELSTEIN LAB

After having characterised the model parameters using the data from Boyle et al. we can without the use of any free-parameters predict the outcome of your experiment.
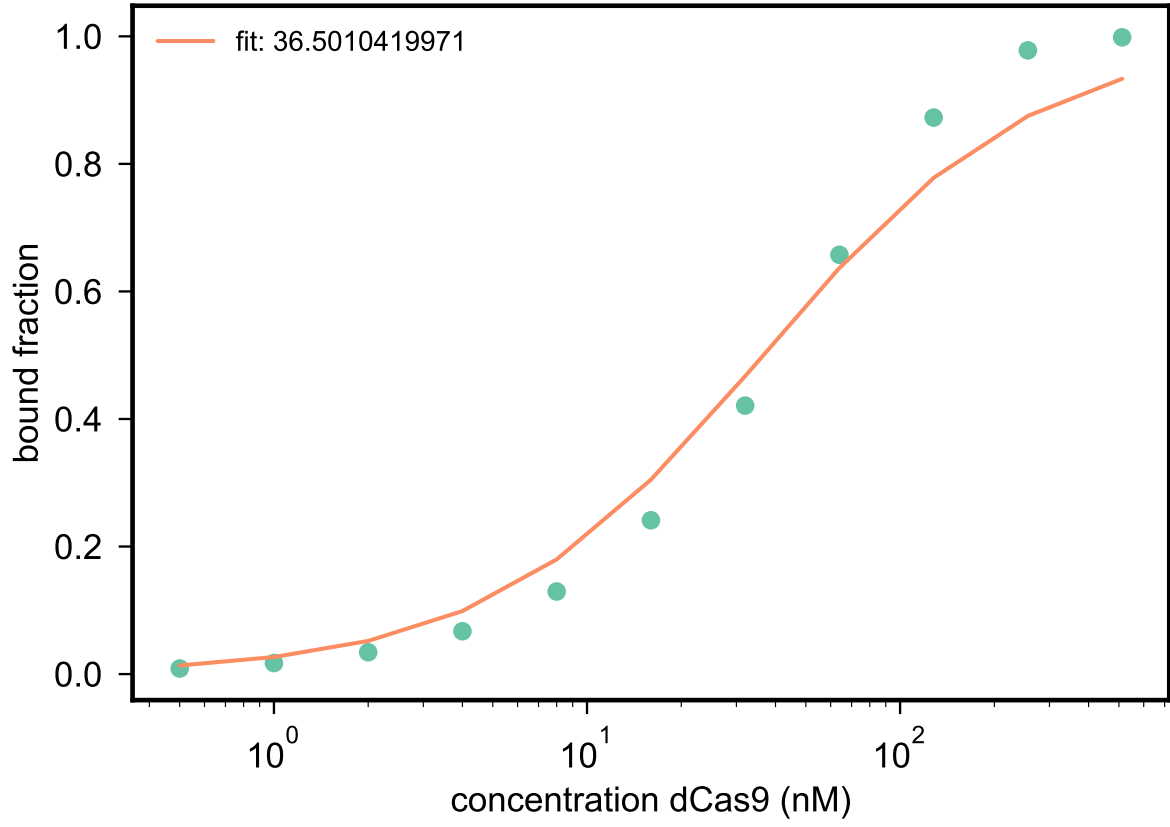
FIG. 6: *We mimic the procedure described in Jung et al. by solving the master equation for the bound fraction after 10 minutes for different concetrations. Just as described, we obtain the ABA by fitting the Hill Eqn to the obtained datapoints to get the dissociation constant and taking its logarithm.*
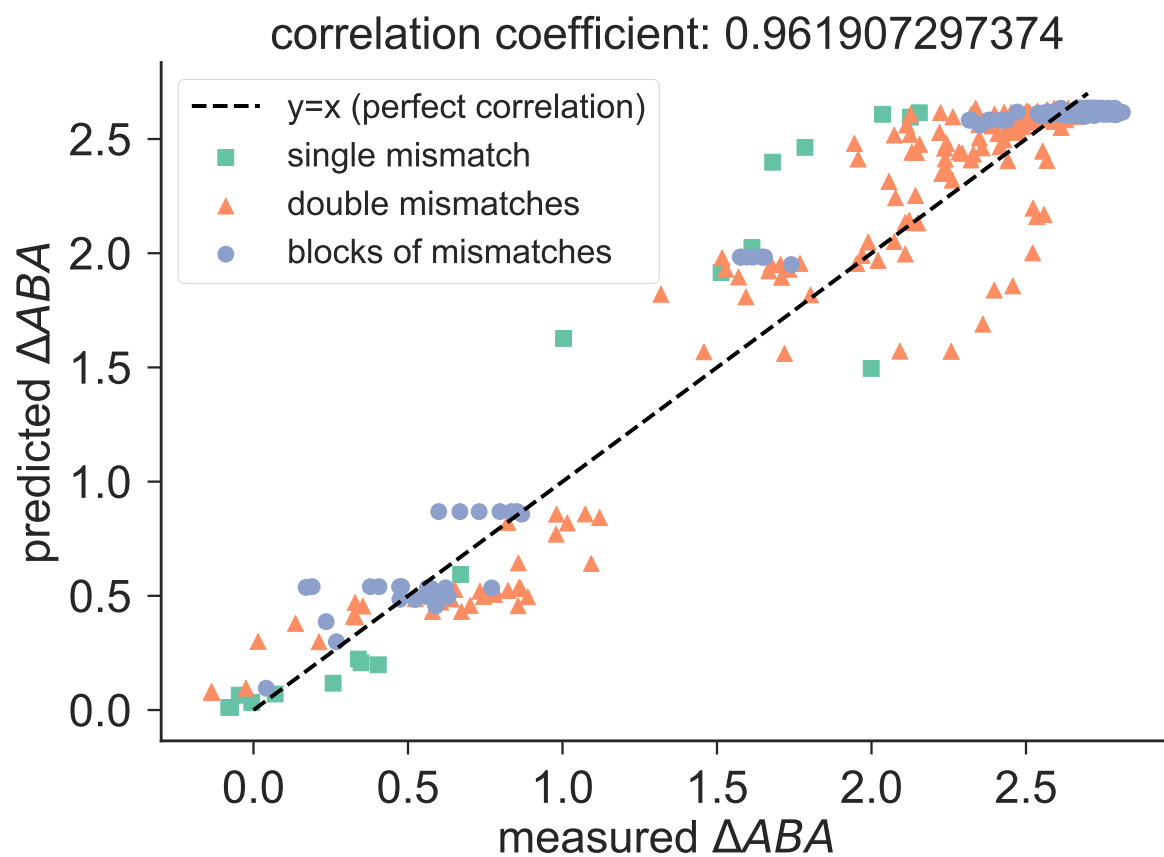
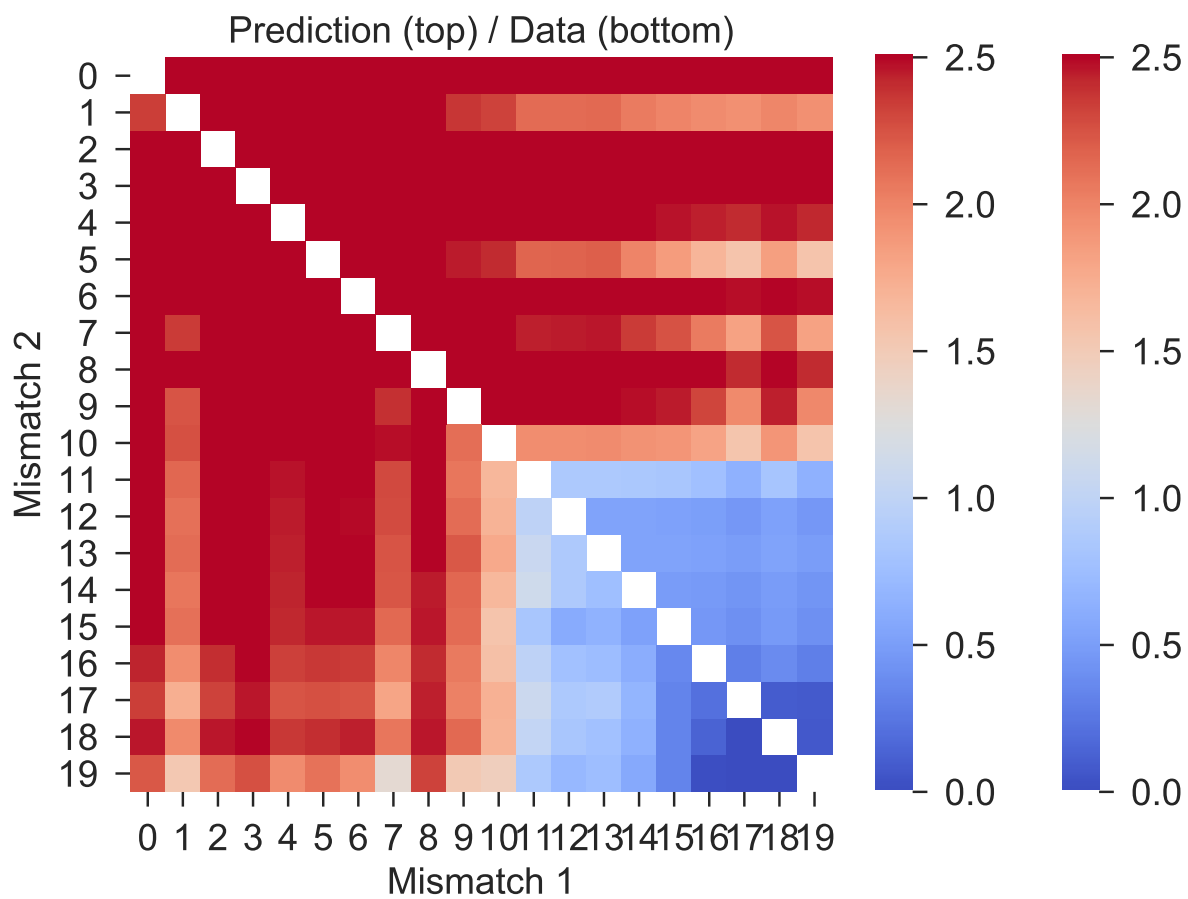**FIG. 7:** *Model prediction versus your data.*
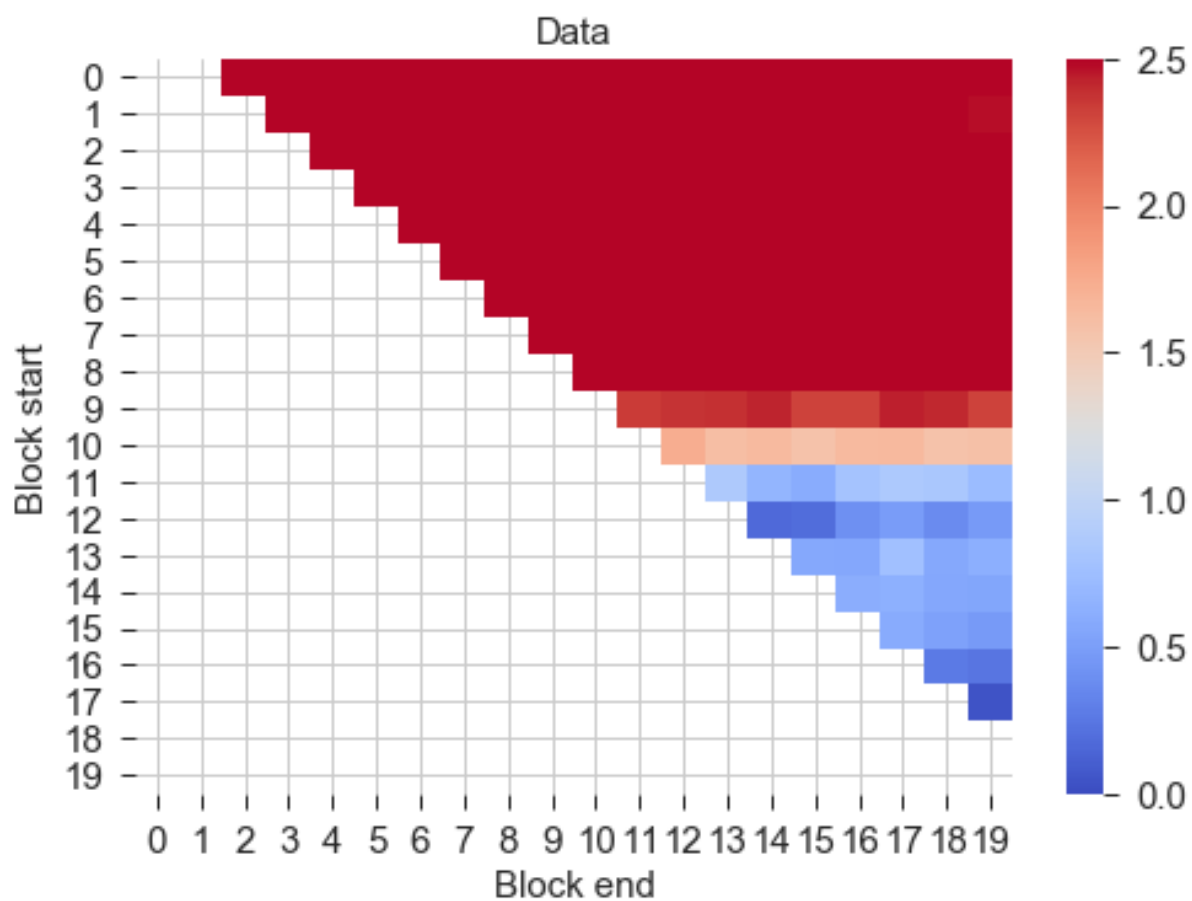
FIG. 8: *all double mismatches*
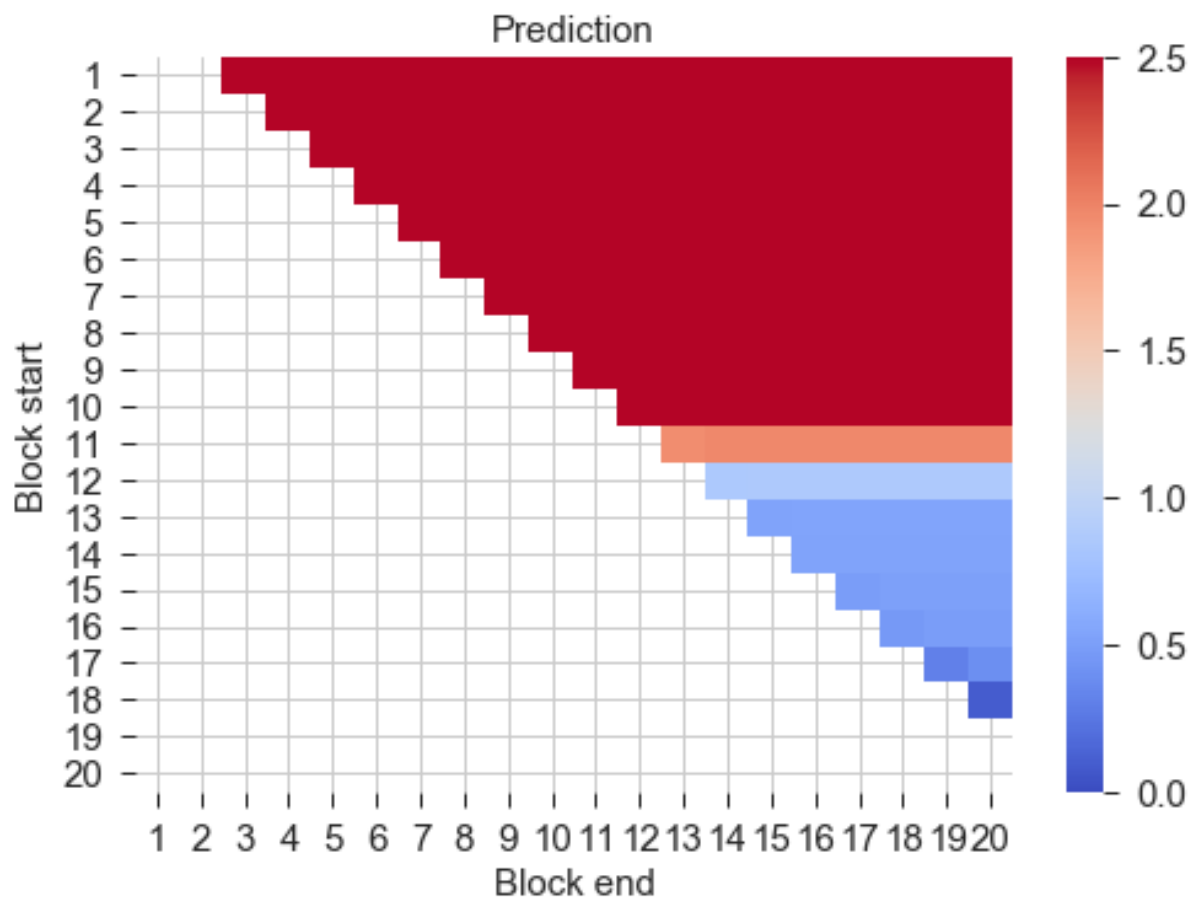
FIG. 9: *data for all sets of consequetive mismatches*

FIG. 10: *prediction of* ΔABA *for consequetive mismatches*
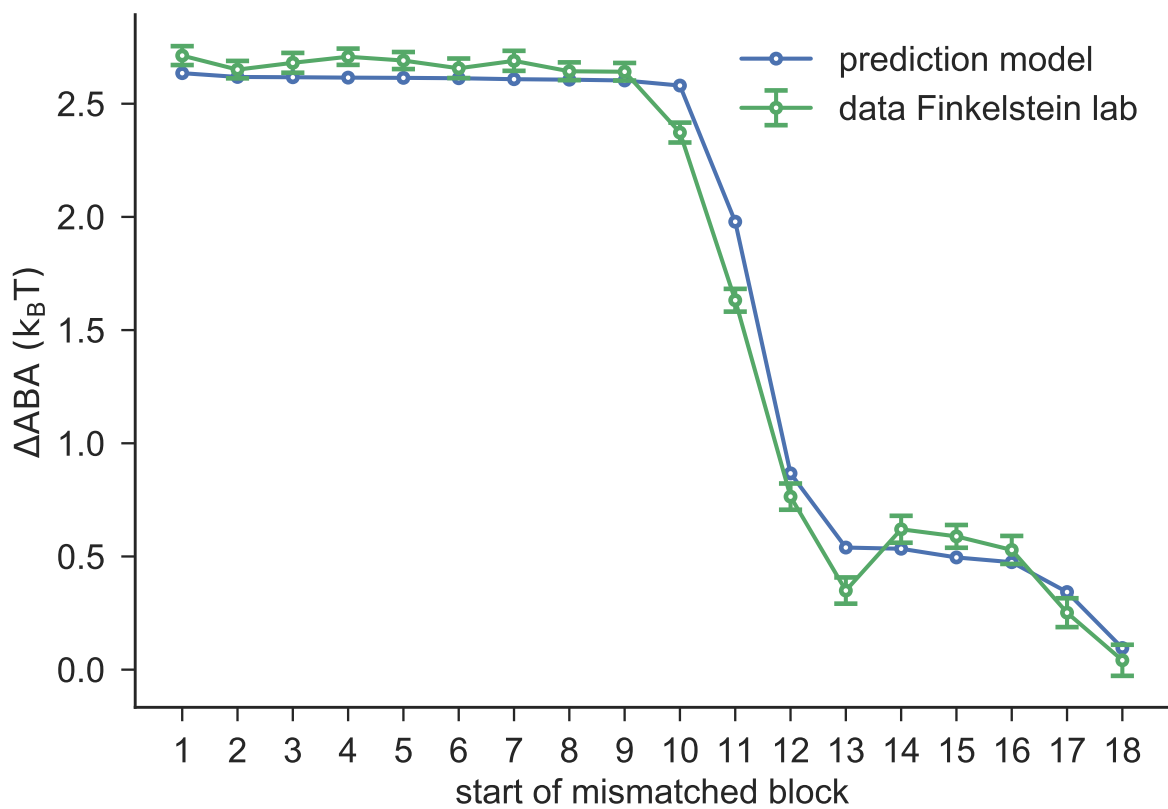
FIG. 11: *Data and prediction for consequetive mismatches after averaging over the block length.*

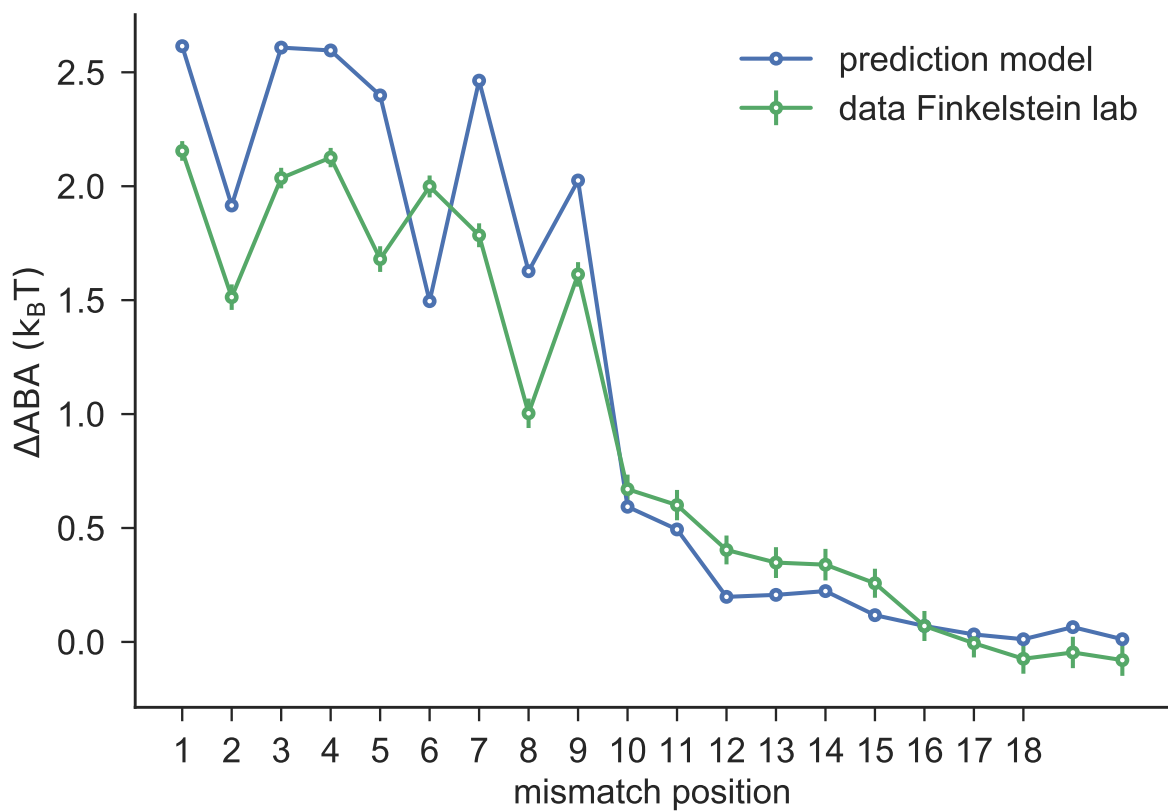FIG. 12: *data and prediction of* ΔABA *for single mismatches.*
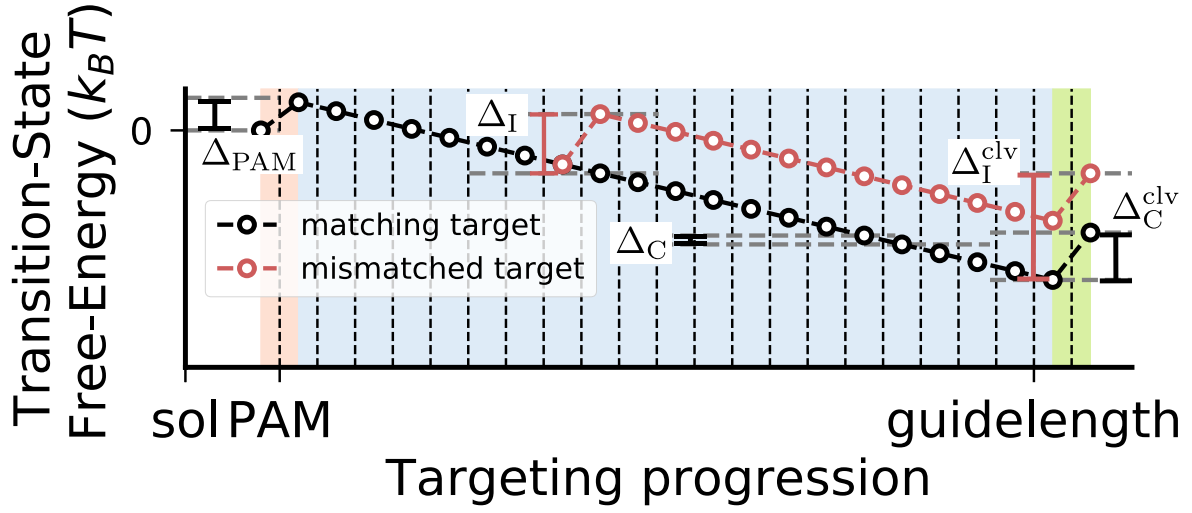
Translating between Cas and (d)Cas:

FIG. 13: *microscopic free-energy landscape for the active enzyme. Parameters are 20 gains/penalties for matches ϵ_C, 1 energy of the PAM bound state ϵ_{PAM}, 20 mismatches penalties ϵ_I and 3 distinct-valued forward rates (binding rate from solution to PAM, the rate from PAM to enter the R-loop and the rate of adding a basepair to the R-loop). Given this is the landscape for the active enzyme, there is another rate in play: the intrisic rate to cleave (or the associated forward barrier).*