Proceedings of the
46th IEEE Conference on Decision and Control
New Orleans, LA, USA, Dec. 12-14, 2007

ThB01.1

Dynamics and Control of Biomolecules: Research Venues and Opportunities

Tutorial Session at the 2007 IEEE Conference on Decision and Control

William B. Dunbar[*,1] Noah A. Wilson[1] Joseph Schaeffer[2] Eric Klavins[3] Joshua Bishop[3] Bertrand C.W. Tanner[4]

Abstract—The purpose of this session is to present to the CDC audience a set of molecular-level biological research venues in which dynamics and control is, or can be, of use. With the large and growing number of people within the control community interested in biology, experiment driven research at the interface of biology and control is enticing. In our session, four different research venues are presented: 1) feedback control of individual DNA molecules in a nanopore, with a goal of DNA sequencing; 2) DNA secondary structure kinetics, with the goal of improving design and synthesis of DNA logic circuits that may be used, for example, for *in vivo* feedback control of biological processes; 3) an artificial DNA nanomotor, for which feedback compensation improves robustness and performance; and, 4) cooperative protein interactions in muscle force generation, in which identification of feedback mechanisms can ultimately improve therapeutic methods for post-cardiac arrest rehabilitation.

Keywords:

I. INTRODUCTION

The purpose of our session is to present to the CDC audience a set of molecular-level biological research venues in which dynamics and control is, or can be, of use. With the large and growing number of people within the control community interested in biology, experiment driven research at the interface of biology and control is enticing, and compliments systems biology research. Four different research venues are presented:

1) **Feedback control of biological polymers in a nanopore**, presented by N.A.Wilson and W.B.Dunbar. This work is collaboration with UCSC Biomolecular Engineering Professors David Deamer and Mark Akeson, staff researchers Seico Benner and Robin Abu-Shumays.

2) **Kinetics of multiple interacting DNA strands**, presented by J. Schaeffer. This work is in collaboration with Prof. Erik Winfree, Departments of Computer Science and Computation and Neural Systems, California Institute of Technology.

3) **Motor control at the nanoscale**, presented by E. Klavins and J. Bishop. University of Washington, Seattle.

4) **Dynamics and feedback in muscle force generation**, presented by B.C.W. Tanner. This work is in

collaboration with Biology Professor T.L. Daniel and Bioengineering Professor M. Regnier, University of Washington, Seattle.

The following common themes will be incorporated, to varying degrees, into each presentation:

- Details on the biomolecular experiement, including motivation.
- Relevance of feedback, as it occurs naturally or is used to engineer autonomy in some respect.
- Details on mechanisms for sensing and actuation, and the limitations of each at the biomolecular level.
- Future theoretical and/or applied research that incorporates the use of tools from dynamics and control.

This paper devotes a section for each of the four presentation topics.

II. FEEDBACK CONTROL OF BIOLOGICAL POLYMERS IN A NANOPORE

This section demonstrates feedback voltage control of individual DNA hairpin molecules captured in a nanopore, summarizing the results in [11]. A finite state machine exectues voltage control logic on a field-programmable gate array for rapid detection and regulation of hundreds of DNA hairpins, one at a time. The voltage control method is an integral part of ongoing research for direct monitoring and control of enzymes bound to biopolymers [4].

A. Introduction

Nanopore sequencing is based on electrophoretically driving a singe-stranded DNA (ssDNA) or RNA molecule through a nano-scale pore [6]. The potential of this technology is high-speed, high throughout sequential identification of all nucleotides in any single DNA or RNA molecule. Many research groups are exploring and developing biological and solid-state nanopores to achieve low-cost, high throughput nanopore-based sequencing [9], in addition to other single molecule sensing applications [7].

In the biological nanopore setup, a planar lipid bilayer is created across a 20 μm teflon aperture in a KCl solution. A single α-hemolysin protein channel is inserted into the planar lipid (Figure 1). The channel (pore) is 15 nm in length and varies in diameter. The cis-opening of the pore is 2.6 nm wide, opening to a 3.6 nm vestibule before narrowing to a limiting 1.5 nm width at the beginning of the stem. The remainder of the stem up to the trans-opening is 2 nm wide. The vestibule is large enough for double-stranded DNA (dsDNA) to enter, but the limiting stem is just wide enough for ssDNA to pass through. Across the bilayer, AgCl
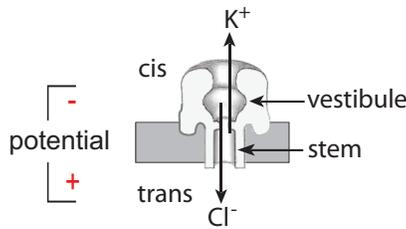
Fig. 1. Graphic of single α-hemolysin protein (mushroom shape) channel inserted into lipid bilayer. Under applied potential trans-side positive, K^+ ions flow to the cis side, and Cl^- ions flow to the trans side. Location of vestibule and stem of pore channel are shown.

electrodes are used to apply a potential that produces an ionic current through the pore. The field created by this voltage pulls the negatively charged phosphate backbone of the ssDNA or RNA through the pore, passing from the cis side to the trans side of the pore with the trans-side voltage positive. Translocating molecules partially obstruct the pore, causing an momentary drop in ionic current. These translocation events are characterized by the amplitude and duration (dwell time) of the attenuated current.

We use DNA oligomer that is 79 nucleotides total in length, with a 20 base pair hairpin (20 bphp). The hairpin is formed by the 3' end folding over and annealing on itself resulting in a 20 base pair region. The hairpin is thus the double-stranded segment, with the single-stranded segment 35 nucleotides long (4 unpaired bases in the doubled-stranded end loop). Upon capture of the ssDNA end, the hairpin enters the pore vestibule and remains until the hairpin is unzipped by voltage force. A schematic of the nanopore system and an example 20 bphp translocation event is illustrated in Figure 2.
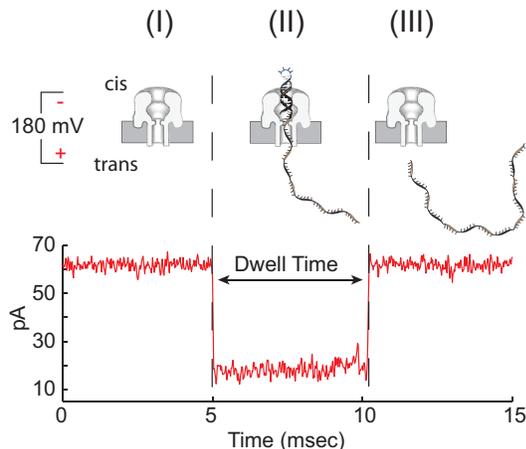

Fig. 2. Schematic of nanopore and DNA, and plot of representative ionic current signal during a 20 bphp DNA translocation event under 180 mV applied potential. (I) At 180 mV, KCl ions pass through the open channel resulting in ~64 pA current. (II) Upon capture of the single-stranded end of the DNA molecule into the cis opening of the pore, the flow of ions is reduced to ~20 pA. (III) After ~5 ms, the voltage unzips the hairpin, causing ssDNA to pass through the pore into the trans chamber, completing the measured blockaded event. The duration of the event is referred to as dwell time.

Regarding the resolution limits of ionic current measurements, homopolymers of ssDNA and block copolymers of RNA are distinguishable based on the measurable differences in the blockade current amplitude or kinetics [2]. However, translocation rates are too fast (up to 2 nucleotides/μsec, [2]) to identify individual nucleotides in heterogeneous single-stranded polymers using existing biological nanopores [7]. In this paper and other studies [8], [10], DNA with single and double stranded segments is used to increase the dwell time of nucleotides in the pore (0.5–10 ms, depending on applied voltage and dsDNA segment length). Use of DNA-binding proteins (enzymes) also increases nucleotide dwell time in the pore. In [4], the interaction of DNA with the Klenow fragment (KF) of Escherichia coli DNA polymerase I was explored. As an initial effort at voltage control of enzyme-bound DNA, we demonstrated efficient (< 3 ms) automated detection of individual ternary (DNA-KF-dNTP) complexes based on a characteristic amplitude, and truncation of the blockade time by voltage reversal. In [11], we demonstrated automated detection and manipulation of DNA hairpins. This paper summarizes the results in [11], and provides new and refined analysis of the results. Direct control of ssDNA in a nanopore has been demonstrated by others [3], in which detection of DNA is based on monitoring the raw amplitude relative to a threshold level. In contrast, we use a windowed mean filter and detection depends on the mean remaining within a preset amplitude range (< 6 pA in spread) for multiple consecutive comparisons. Alternative methods for single molecule sensing and manipulation include optical tweezers [1] and atomic force microscopy [5]. At present, greater spatial and temporal resolution of single DNA molecule polymerization has been achieved with optical trapping methods than with nanopores. However, optical trapping methods generally require more preparative steps, and far fewer molecules can be analyzed over a common time period.

B. Equipment and FSM Control Logic

The nanopore system is setup in a 0.3 mM KCl solution. A patch-clamp amplifier, Molecular Devices AxoPatch 200B, regulates the applied voltage and measures the ionic current through the channel. The data are recorded using the Molecular Devices Digidata 1440A digitizer, sampled at 50 kHz and low-pass filtered at 5 kHz with a four-pole Bessel filter. The voltage control logic is programmed using a finite state machine (FSM) within LabVIEW 8 software. The FSM logic is implemented on a field-programmable gate array (FPGA) hardware, National Instruments PCI-7831R. An FPGA is a reconfigurable hardware platform that permits fast measurement and voltage reaction times (1 μsec output sample time). A mean filter applied to the incoming current signal on the FPGA removes a large portion of the peak-to-peak noise. Specifically, every 5.3 μsec, the FPGA samples the ionic current and computes a windowed mean amplitude based on the previous 0.75 ms of signal. Every 0.2 ms, the FPGA tests if the mean is within 20 ± 2.8 pA (17.2 to 22.8 pA range). The basis for choosing this range is that ~20

3298

pA is the median amplitude for DNA 20 base pair hairpin events at 180 mV, as shown in the experimental results below. If the mean enters and remains within this range for four consecutive tests, the FSM logic diagnoses the blockade as a DNA hairpin event. The *nominal detection time*, between DNA translocation event and diagnosis of the event, is 2.0 ms; 0.75 ms for the windowed mean to first enter the 17.2 to 22.8 pA range, and 0.6 ms for three more confirmed tests, and 0.65 ms of delay[1].

C. Experiments and Results

In our first experiment, the objective was to efficiently detect individual DNA hairpin events, and increase the blockade dwell time by lowering the applied voltage from 180 mV to 150 mV upon detection. This is referred to as **dwell time extension control**. Next, we sought to aggregate the extended blockade dwell times, by expelling the DNA using voltage reversal of -50 mV after 10 ms at 150 mV. This is referred to as **dwell time aggregation control**. The motivation was to increase the nominal hairpin dwell time, and expel the molecule before unzipping the hairpin. A typical 20 bphp event at constant 180 mV voltage is shown in Figures 2 and 3aI. The probability histogram of the base 10 logarithm of dwell time (Figure 3aIII, blue) is unimodal, with median dwell time of 2.8 ms. The median amplitude of the event plot in Figure 3aII is 20.9 pA with an interquartile range (IQR) of 1.7 pA. Only 6% of events are in the subset range of 13-18 pA (3aIII, yellow). For the same experiment at constant 150 mV voltage (data not shown), the events cluster around a median amplitude of 15 pA and 87% of 150 events are in the 13-18 pA range. Thus, under extension and aggregation control for which the voltage is reduced to 150 mV for all detected events, a larger percentage of blockades should have a mean amplitude within the 13-18 pA range.

1) Dwell Time Extension Control (Figure 3b): Upon diagnosis of a DNA hairpin event using the mean filtered current, the command voltage is reduced to 150 mV until the hairpin unzips and the DNA translocates through the pore. Using 180 mV for capture results in more events than 150 mV, while reducing to 150 mV extends the life of the hairpin. Dwell time extension is useful for terminal base-pair sequencing by machine learning methods [10]. After each translocation, the FPGA resets the voltage to 180 mV. A representative event is shown in Figure 3bI. The event plot (Figure 3bII) pattern shows that events faster than the nominal diagnosis time of 2.0 ms are unaffected by extension control, and events with longer dwell times converge to the ∼15 pA mean amplitude as expected. The concave trend is also consistent with an equation for event's mean amplitude vs. dwell time. In particular, for an event at 21 pA (median amplitude at 180

mV) for 2.4 ms [2], and at 15 pA (median amplitude at 150 mV) for x ms, an approximate mean amplitude $\bar{I}$ is

$$\bar{I} = \frac{2.4 * 21 + 15 * x}{2.4 + x}. \qquad (1)$$

When $x \approx 24$ ms, as in Figure 3bI, $\bar{I} = 16$ pA. Equation (1) closely matches the mean amplitude vs. dwell time data (Fig. 3bII). Also, the fraction of events within the subset range 13-18 pA increased to 41%, as shown in the yellow histogram overlaid on the blue probability histogram (Fig. 3bIII).

2) Dwell Time Aggregation Control (Figure 3c): The objective was to aggregate the dwell times of the extended events by applying 150 mV for 10 ms upon diagnosis of a hairpin event, followed by voltage reversal of -50 mV for 5 ms. The reversal time of 5 ms is known to be sufficient to clear the DNA from the channel, prepping the pore for the next event. Aggregation control would imply a measure of control over the distribution of the events, in addition to temporal control of individual molecular events. A representative event is shown in Figure 3cI. As before, the event plot (Fig. 3cII) pattern shows that events faster than the nominal diagnosis time of 2.0 ms are unaffected by aggregation control. Within the subset range of 13-18 pA, the median amplitude is 16 pA with 0.7 pA IQR (amplitude histogram shown in Fig. 3cII). The 16 pA median is consistent with (1), since for $x = 10.0$ ms, $\bar{I} = 16$ pA. Also in the subset range, and the median dwell time is 12.4 ms with 0.1 ms IQR. The low IQR indicates a high degree of control over the distribution of events that extend to at least 10 ms at 150 mV. The median dwell time of 12.4 ms is commensurate with 2.0 ms of detection time, 10 ms at 150 mV, and 0.4 ms due to a transient that is included at the end of each event resulting from voltage reversal[3]. Summary statistics for the histograms in Figure 3III are reported in Table I.

TABLE I
SUMMARY STATISTICS FOR FIGURE 3III.

Figure No.	No. of Events	Median Dwell Time (ms)	IQR (ms)
2aIII[a]	472[b]	2.8	4.2
2bIII[c]	76[d]	31.6	62.0
2cIII[c]	256[e]	12.4	0.1

[a]Blue histogram, for events within 10 to 30 pA range.
[b]6% (27 events) within subset 13-18 pA range.
[c]Yellow subset histogram, for events within 13-18 pA range.
[d]41% of the 187 events within 10 to 30 pA range.
[e]55% of the 466 events within 10 to 30 pA range.

[1]Certain inefficiencies in FPGA signal routing into the sampling loop caused the additional 0.65 ms of delay in the reaction time. By bringing global signals inside the sampling loop, the delay has recently been eliminated, reducing detection time to 1.35 ms.

[2]Step changes in voltage induce a capacitive transient, and the transient at the end of each event is ∼0.4 ms for changing from 150 mV to 180 mV. Thus, 2.4 ms at 21 pA is 2.0 ms of detection time and 0.4 ms of transient time. While the 0.4 ms transient varies in amplitude, assuming 21 pA is sufficient for line fitting.

[3]The transient due to the 180 mV to 150 mV change is included within the 10 ms waiting time under aggregation control.
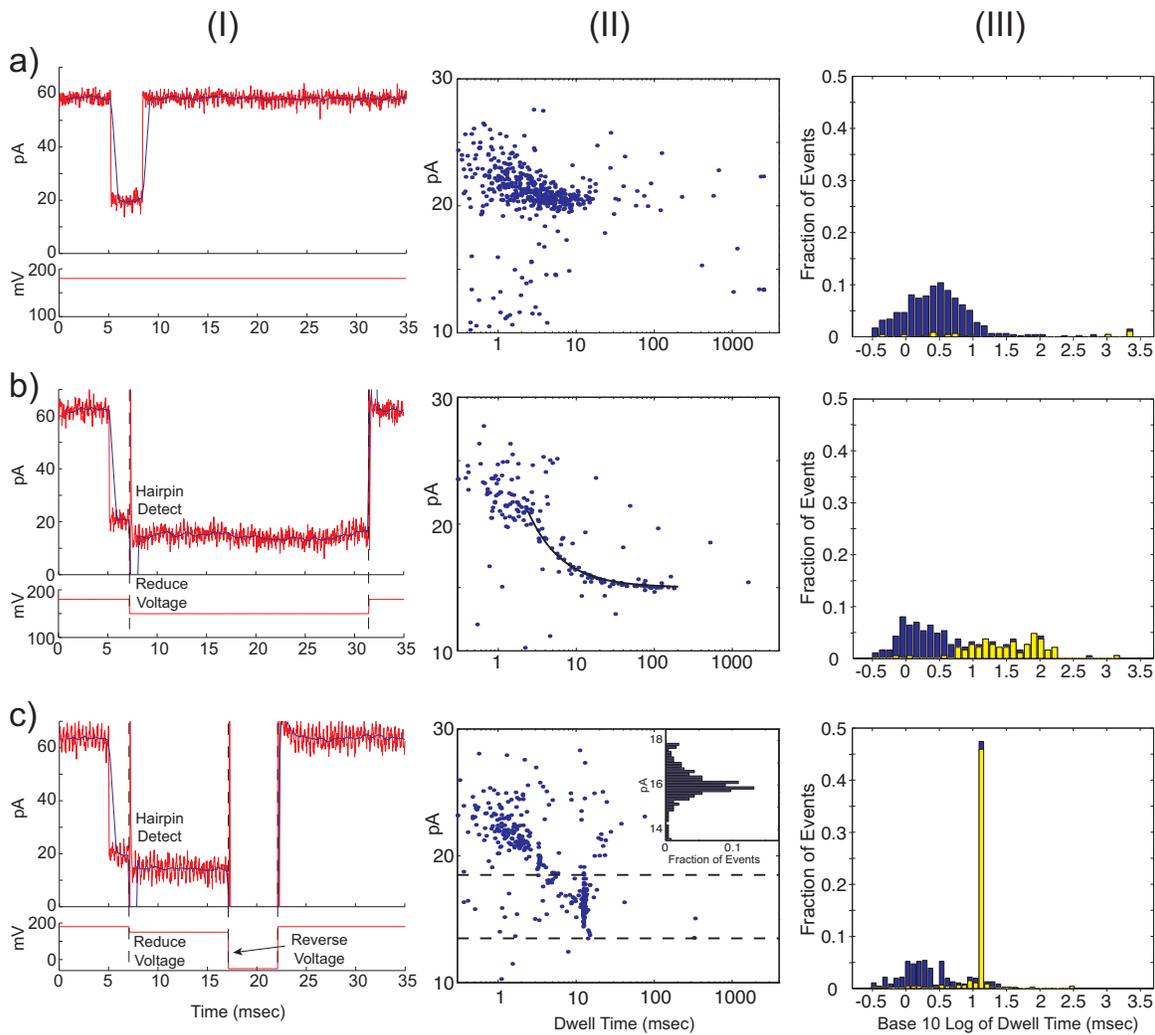
Fig. 3.   Regulation of 20 bphp dwell time using FSM control. (I) The red current signals are low-pass filtered at 5kHz, the blue signal is a mean filtered current, and the red voltage signal is the commanded voltage. Typical events and corresponding voltage signals under a) constant 180 mV voltage, b) dwell time extension control, and c) dwell time aggregation control. (II) Event plot of DNA events, showing average amplitude vs. dwell time for each event (point). Equation (1) (line) fit to events in bII), and amplitude histogram for events within 13-18 pA (dashed line) range in cII). (III) Probability histograms of the base 10 logarithm of dwell time for all events (blue), and for subset of events in 13-18 pA range (yellow).

D. Conclusions

This section demonstrates our control authority to extend and aggregate the dwell times of hundreds of DNA hairpin blockade events in a nanopore. Additionally, it complements our recent work [4], confirming our ability to detect DNA-enzyme blockades and DNA hairpin blockades. Ultimately, nanopore-based characterization of enzyme dynamics will require direct detection and control of multiple DNA conformations (states) relative to the enzyme, and direct control of enzyme-free DNA is a prerequisite toward developing this capability.

III. KINETICS OF MULTIPLE INTERACTING DNA STRANDS

This section examines the modeling and simulation of DNA secondary structure kinetics, with applications to the prediction and refinement of the DNA logic circuits in [12].

A. Introduction

Within cells, DNA is used to store the genetic code that is required for cells to reproduce and perform their functions. Taken outside of the cell, DNA has been used for many purposes by researchers, such as DNA logic gates [12]. DNA utilized as a gate component can receive single stranded DNA inputs and produce a DNA output based on the logic function the gate implements. These gates can be composed into large circuits analogous to electronic circuits, and in theory similar DNA circuits could be placed back into the cell, to perform diagnostic or regulatory functions. The kinetic characteristics of these gates are important to analyze as they allow us to determine whether the gate is correct and usable in a larger circuit. Simulating the structure kinetics allows us to predict these characteristics and refine gate designs.

In order to simulate these systems, we must decide the level of detail to be simulated. The secondary structure of

3300

a DNA molecule is the hydrogen bonding (called base-pairing) information within the molecule: the guanine bases that are paired with cytosine bases and the adenine bases that are paired with thymine bases. The tertiary structure of a DNA molecule is the full three dimensional atomic coordinates for the molecule. The simulation of tertiary structure requires a large amount of computational power to simulate the molecular dynamics over very small time frames. Since millisecond to hour timescales are needed to ensure DNA circuit logic correctness, a coarser simulation of secondary structure is investigated in this work to bypass the computational demands of tertiary structure simulation.
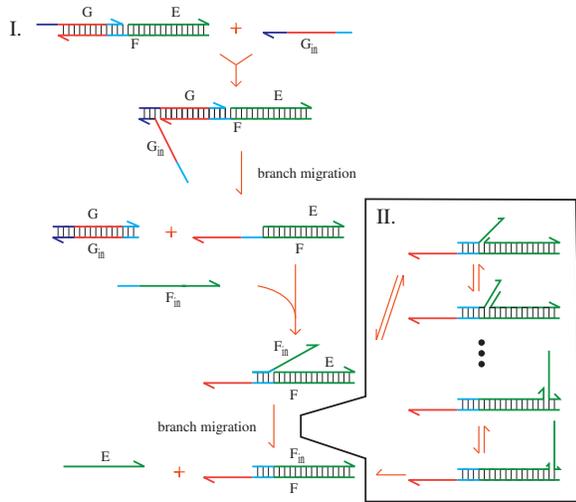

Fig. 4. (I) Schematic for DNA-based AND logic gate. The 3' end of a strand is designated by an arrow. Sequences appearing in the same color are equivalent or complementary, as appropriate. (II) Branch migration step in a logic gate.

A simple DNA logic gate representing a two input AND gate is shown in figure 4, along with the reaction pathway it is expected to follow. Inputs to the AND gate are represented by single unpaired DNA strands, and the gate is a complex of several DNA strands. The reaction proceeds by having the first input bind to a small sequence on the gate molecule and then displacing the strand previously bound to that part of the gate, by a process known as branch migration. This forms a waste product which disassociates from the remaining part of the gate, leaving the gate accessible to the second input. The same process occurs with the second input, and we are left with two waste molecules and a single stranded output from the gate. This output could then be used as the input to another gate, and in this manner enables us to build larger circuits out of these basic components.

The design and characterization of these components is essential for composing larger circuits. Choosing the sequences used for these gates such that they follow the intended pathways is known as the sequence design problem. Using arbitrary choices for the sequences can lead to many issues with the resulting system, such as steps in the pathway being extraordinarily slow or having strands that are intended to be unpaired actually having a large amount of secondary

structure. More intelligent approaches to sequence design must take into account both thermodynamic and kinetic considerations to produce a viable system.

B. Energy Model

The conformation of a nucleic acid strand at equilibrium can be predicted by a well studied model [14] [15]. The distribution of states at equilibrium should follow a Boltzmann distribution, where the probability of observing the strand in state i is given by

$$Pr(i) = e^{-\Delta G(i)/RT}. \qquad (2)$$

Determining the free energy of the state i, $\Delta G(i)$ is the key feature of these energy models. Recent work has extended the basic model to cover systems with multiple interacting nucleic acid strands [16]. This revised energy model forms the basis for our simulation.

The energy model is based on nearest neighbor interactions between the nucleic acid bases. These nearest neighbor interactions divide the secondary structure of the system into local components which we refer to as *loops*. The energy of each loop is then independent of the rest of the secondary structure, and we can break down these loops into various categories as shown in figure 5. Parameter tables for each category have been determined from experimental data.
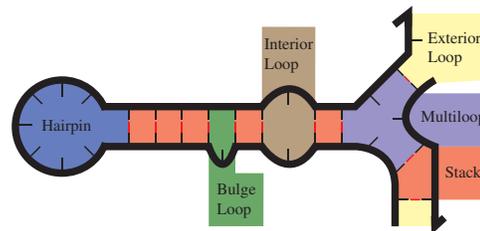

Fig. 5. Loop decomposition of a secondary structure.

Computation of $\Delta G(i)$ from these loops can be done by summing over the contribution from each component. Each complex also includes a strand association penalty $\Delta G_{assoc}(N-1)$, where N is the number of strands in the complex, which accounts for the entropy loss of the system from including multiple strands in the complex.

The energy of a state j of the system can then be broken down based on the complexes k in the state:

$$\Delta G_{sys}(j) = \Sigma_{k \in j} \Delta G_{complex}(k) \qquad (3)$$

Then the energy of each complex k is just the sum over all the loops l in the complex, plus the association term:

$$\Delta G_{complex}(k) = \Delta G_{assoc}(N-1) + \Sigma_{l \in k} \Delta G_{loop}(l) \quad (4)$$

C. Kinetics Model

Thermodynamic predictions have only limited use for some systems of interest, as the key information to be gathered is the reaction rates and not the equilibrium states. The gate system mentioned above has a well defined ending state we can find by thermodynamic prediction, but predicting whether it will reach this completed state in a reasonable

3301

amount of time requires modeling the kinetics. Kinetic analysis can also help uncover poor sequence designs, such as those with alternate reactions leading to the same states, or kinetic traps which prevent an intended reaction from occurring quickly.
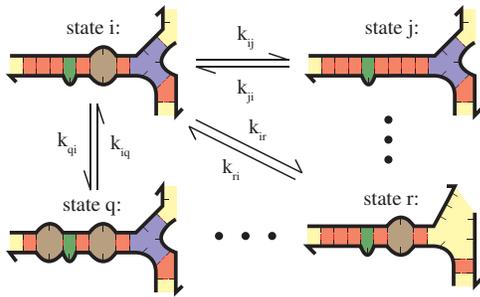

Fig. 6. Diagram of a current state i and some adjacent states, with transition rates.

The kinetics are modeled as a continuous time Markov process over secondary structure space. States are considered adjacent if they differ by a single base pair (Figure 6), and we choose transition rates such that we obey detailed balance:

$$\frac{k_{ij}}{k_{ji}} = e^{-\frac{\Delta G(j) - \Delta G(i)}{RT}} \qquad (5)$$

This property ensures that given sufficient time we will arrive at the same equilibrium state distribution as the thermodynamic prediction. Given that we are currently in state i, the next state m in a simulated trajectory is chosen randomly among the adjacent states j, weighted by the rate of transition to each.

$$Pr(m) = \frac{k_{im}}{\Sigma_j k_{ij}} \qquad (6)$$

Similarly, the time taken to transition to the next state is chosen randomly from an exponential distribution with rate parameter equal to the total rate out of the current state.

$$Pr(\Delta t) = exp(-\Sigma_j k_{ij} \Delta t) \qquad (7)$$

D. Algorithm Efficiency

Previous work implemented similar models for single strands [13], but in order to handle larger systems like DNA gates we had to solve two problems: how to handle systems with multiple strands, and how to compute kinetic trajectories efficiently. The first problem is overcome by a new energy model that provides the necessary energy computation for multistranded systems, as well as the addition of kinetic moves which can join or break complexes. Computation of a single trajectory was made more efficient by taking advantage of several local properties inherent in non-pseudoknotted secondary structure. Energy computation for a system depends on the energies of all the individual loops within the secondary structure, and these are independently computable, so that when one facet of the secondary structure changes, only a small portion of the energy contribution must be recalculated.

Generating all the possible kinetic moves from a given state is also made more efficient by taking advantage of local structure. A single move can only affect the possible moves available to the affected loop(s) and their immediately adjacent loops, and so by storing the lists of moves locally we can avoid regenerating the entire move list for the whole structure at each kinetic step.

E. Simulation of a DNA gate

The experimental fluorescence data represents the percentage of all the gate molecules which have produced an output, as a function of time. The gate system from figure 4 is used with two additions: a fluorophore is added to F at the 5' end, and a quencher is added to E at the 3' end. This makes the final step, where the F strand disassociates from the E strand, lead to an increase in fluorescence. This data appears in figure 7I for the given system.

Data from our simulation takes a subtly different form. Each simulated run of the system is a trajectory in system state space which represents a single conformational path the molecule(s) could have taken. By measuring the first passage time of the final step of the system (the first time the E strand disassociates from the molecule), we get data that can be accumulated over many trajectories. This produces a graph of the percentage of completed trajectories over time, which can give us similar information to the fluorescence experiments. This data is shown in figure 7II for the same set of initial conditions as the experimental data.
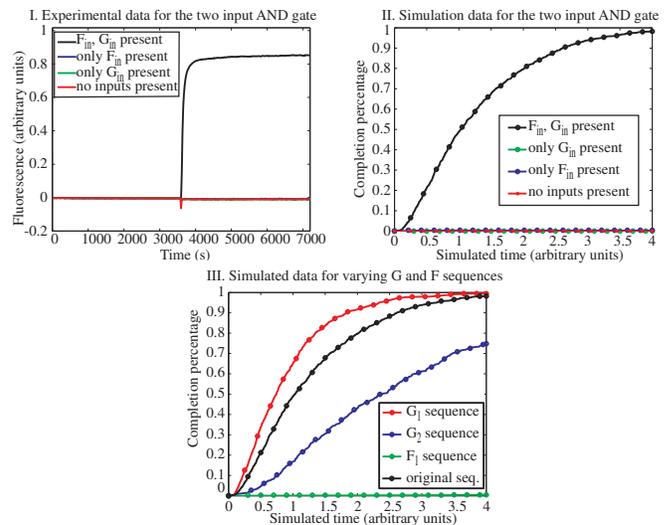

Fig. 7. (I) Fluorescence data from experiment [12], with varying inputs adding one hour into experiment. (II) Simulated data for the percentage of completed trajectories over time, with varying inputs. (III) Simulated data with varying sequence designs for the gate system.

Comparing the simulated data, we see that it predicts the same qualitative behavior as observed in the experimental system: when both inputs are present, it reaches a fully on output quickly, and when any input is missing there is very little leak in the same time frame. Furthermore, we can test many different sequence designs with the simulator and rank them, shown in figure 7III. For example, we tried

3302

two alternate sequences for the G strand (also changing complementary and equivalent sequences as appropriate), altering only those bases which are exposed during the initial step. Simulations predicted that the same reaction with both inputs present would be faster for the G_1 sequence, and slower for the G_2 sequence. We also tried an alternate sequence for the F strand with a drastically different result: the modified F_1 sequence is extremely slow in reaching the final state, even though though that state is still the most favorable equilibrium state. Individual trajectories can also be visually examined to diagnose kinetic problems within the system. For example, we can look at the secondary structure conformation in a particular trajectory for the F_1 sequence and by examining the changes in secondary structure along the duration of the trajectory we can pinpoint the secondary structure that causes the pathway to be slow, in this case a strong hairpin blocking branch migration.

F. Conclusions

This section examined the area of DNA secondary structure kinetics, including the models necessary for simulation as well as a system for which prediction of the kinetics is a useful tool. The models and simulation then allows for the possibility of feedback during the sequence design process, using the simulated data to refine our sequences before using them in experimental studies. The study of these DNA gates and circuits has reached the point where circuits involving feedback are possible, such as exponential amplifiers and other components.

As the repertoire of gate components increases, we can begin to imagine placing these DNA circuits into the cell. There they could act as controllers, processing inputs from sensors within the cell, such as aptamers and ribozymes, and sending outputs to cellular actuators, such as flagella. Other possibilities are diagnostic devices, presenting a fluorescent output if certain factors are present in the cell, or even regulatory devices, inhibiting or promoting protein expression based on cell conditions.

IV. MOTOR CONTROL AT THE NANOSCALE

Nanomotors capable of controlled motion at the molecular scale have potential applications in, for example, drug delivery, gene therapy, and nanoscale manufacturing. These applications have motivated recent demonstrations of an assortment of nanomotor designs that use a variety of methods for power, control, and construction [19]–[22].

We are interested in motors that operate autonomously (no external control signal required to actuate the motor) [18], [23], [24]. For example, the DNA nanomotor introduced by Chen, Wang, and Mao [18] is constructed from a 10-23 DNA enzyme [25] (DNAzyme), as shown Fig. 8a. This nanomotor cycles autonomously as the DNAzyme binds to fuel strands and cleaves them into *waste* strands. Unfortunately, the fuel and waste strands have similar affinities for the nanomotor. The result is that the accumulation of waste significantly interferes with the operation of the system, which can be observed experimentally as a loss of performance, as shown in
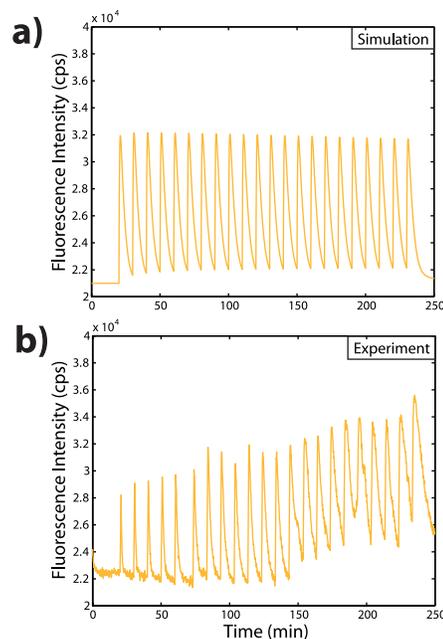

Fig. 9. The compensated system model is the composition of these reactions and those in the original system model depicted in Fig. 8a. The enzyme RNase H degrades fuel and waste bound to motors, mitigating the effect of accumulating waste on nanomotor performance. The model yields a simulated trajectory (a) of the compensated system that is in agreement with experimental results (b). The experimental data show an upward trend not reflected in the simulated trajectory, which may be a result of non-uniform mixing or RNAase H degradation, neither of which are modeled.

Fig. 8c, which represents our reproduction of the experiment by Chen, Wang and Mao [18] (with minor modifications). This problem suggests the need to manage the waste in this system.

In [17], we offer an explanation for the observed behavior of the DNAzyme nanomotor. This explanation suggests a modification of the DNAzyme nanomotor system, which we call the *compensated* system, that incorporates a waste management mechanism and shows improved performance in experiments, Fig. 9. We also introduce a *streamlined* version of the compensated system, Fig. 10, that simplifies the design of the nanomotor. Both the compensated and streamlined systems use the enzyme ribonuclease H [26] (RNase H) to digest waste strands so that they do not interfere with the normal operation of the nanomotor. Finally, we offer an explanation, loosely based on the idea of disturbance rejection, for the improved behavior of the compensated and streamlined systems. Specifically, we show that the original system is highly sensitive to the presence of waste, while the other systems are not, as shown in Fig. 11. This argument is independent of the exact values of the physical parameters in the model [17].

V. COOPERATIVE PROTEIN INTERACTIONS IN MUSCLE FORCE GENERATION

In striated muscle (cardiac and skeletal), the sigmoidal relationship between steady-state force and intracellular $[Ca^{2+}]$ is strong evidence for cooperative interactions among muscle
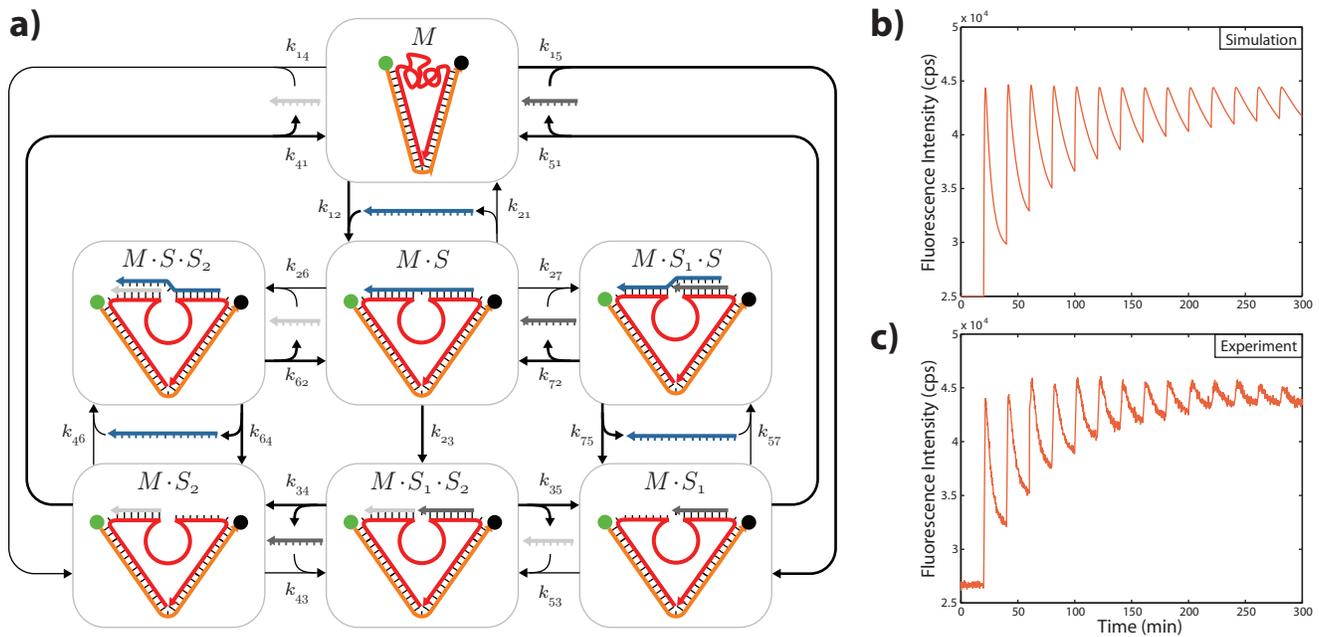
3303

Fig. 8. (a) The reaction network model of the original DNAzyme nanomotor system. The states of the nanomotor are boxed and labeled by species, where M denotes the nanomotor, S denotes the RNA fuel, and S_1 and S_2 denote the waste products. The symbol $(\cdot)$ denotes the hybridization of species into a larger complex. The model captures the operation of the nanomotor as well as the effect of waste having non-negligible affinity for the nanomotor. All reactions are reversible hybridization reactions, excepting the fuel-cleaving reaction, and the direction of these reactions is indicated by bold arrows. Parameters and further details are described in the [17]. The model yields a simulated trajectory (b) of the system that is in agreement with our experimental results (c). In these plots, spikes in fluorescence intensity correspond to the addition of a stoichiometric quantity of RNA fuel to the solution of nanomotors.

proteins. This section demonstrates an integrated experimental and computational platform to examine the influence of architecture in this protein network on force production muscle. Cellular and molecular techniques can alter the spatial distribution of Ca^{2+}-binding regulatory proteins, thereby reducing the number of motor proteins generating force [27]. Because cooperative interactions among muscle proteins are fundamentally coupled (and difficult to separate experimentally), we expand our computational analyses from prior models [28], [29] to separate cooperative pathways in muscle. Muscle is an excellent example of a spatially-complex control system.

A. Introduction

Combining computational and experimental methods [27], [29], [31] allows permits quantifying the relative contributions to force generation from coupled, cooperative protein interactions in cardiac and skeletal muscle. Identifying the relative influence of different cooperative interactions improves our understanding of coupling in protein networks and focuses therapeutic efforts to control debilitating muscle diseases or enhance cardiac function following a heart attack.

Muscle contraction results from interactions between a network of Ca^{2+}-binding proteins that regulate a network of force-generating motor proteins (Figure 12a). These interactions occur in a highly structured lattice where a single protein's activity (such as Ca^{2+} binding or force generation) may influence neighboring protein behavior to amplify force generation (Figure 12b). The sources of cooperativity likely result from allosteric interactions between the proteins actin,

myosin, troponin, and tropomyosin conspiring to produce a large change in force given a small change in intracellular calcium concentration ($[Ca^{2+}]$) [30]. To formulate this control problem, consider the protein interactions leading to contraction: Ca^{2+} binding to troponin, the subsequent movement of tropomyosin to expose myosin binding sites on actin, and binding of myosin with actin (forming myosin cross-bridge; XB) to generate force. Mechanical coupling in the network leads to cooperative protein interactions, creating a positive feedback loop to generate a maximal amount of force for a $[Ca^{2+}]$.

Cooperativity relies upon coupling between neighboring proteins, such as interactions between the thin filament regulatory units (i.e., within the regulatory protein network) or interactions between thick and thin filaments following XB formation (i.e., between these networks) [30]. Figure 12b_i and b_{ii} illustrate how activation of a single regulatory unit (RU; 1 troponin, 1 tropomyosin, and 7 actins) may increase the probability of activation at adjacent RUs (RU-RU cooperativity). RU-RU cooperativity, therefore, amplifies force generation at the molecular level by mechanically translating the activation signal along the regulatory protein network (via overlapping tropomyosin proteins). Another potential cooperative protein interaction augments activation throughout the regulatory protein network following XB binding (XB-RU cooperativity). In both of these examples, the amplification of force given a small change in $[Ca^{2+}]$ relies upon coupled interactions between neighboring proteins in a the network. The degree to which architecture (or
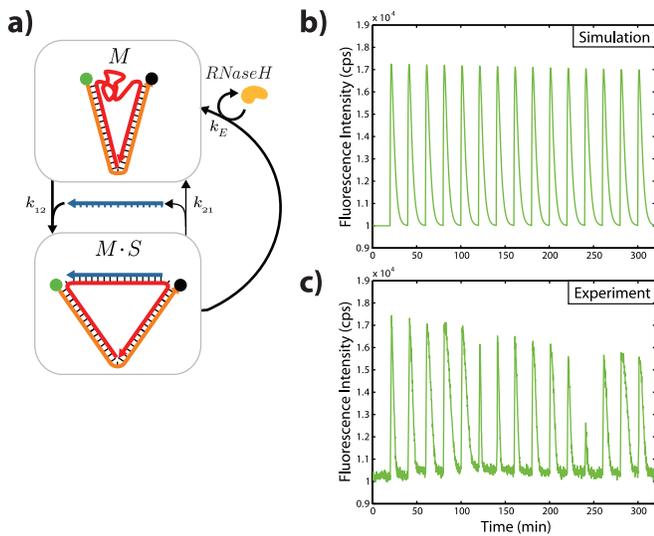
a)

b)

c)

Fig. 10. (a) Diagram of the reaction network model of the streamlined system. The simplified nanomotor design yields a model with only two nanomotor states: *open* and *closed*. The motor opens when it binds with an RNA fuel strand, and closes when the bound fuel is digested by RNase H. The model yields a simulated trajectory (b) of the streamlined system that is in good agreement with experimental results (c). In experiments, RNase H is replenished with every fifth addition of fuel to counteract the natural decay of the enzyme at $22^{o}C$.
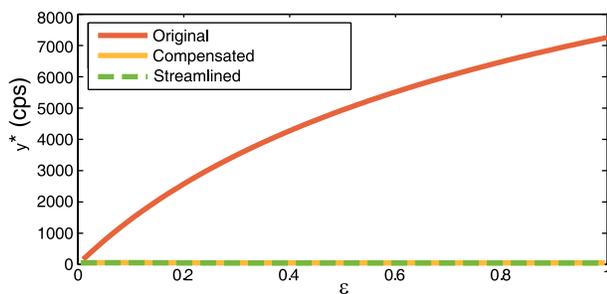

Fig. 11. Plot of the sensitivity of nanomotor performance to the effect of competition from waste strands. the parameter ϵ determines the strength of the competion (details in [17]). The performance metric is a measure of the difference between the base fluorescence level and the equilibrium fluorescence after one addition of a stoichiometric quantity of fuel.

spatial coupling) of this protein network influences cooperative protein interactions becomes an important component regulating muscle performance.

B. Methods

The bulk of experimental studies in striated muscle examine force responses following changes in $[Ca^{2+}]$ or [ATP] (which drives the actomyosin reaction) and its products [ADP] and phosphate. Chemically skinning muscle cells allows full control of intracellular solution conditions and facilitates exchanging regulatory proteins along thin filaments in single muscle cells. These techniques have significantly increased our understanding about cooperative protein interactions in skeletal [27] and cardiac [31] muscle by separating or eliminating some cooperative interactions while monitoring others. While these techniques allow us to alter the spatial or kinetic components of the regulatory
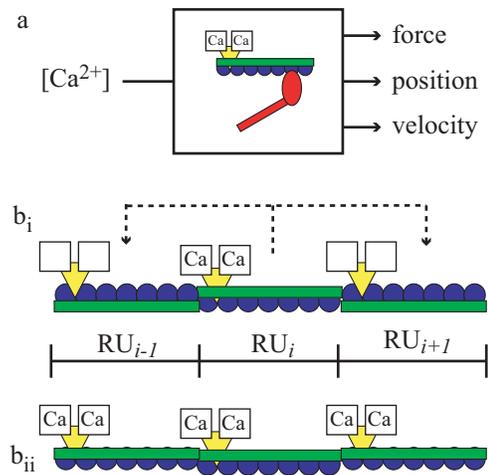
a

b_i

b_{ii}

Fig. 12. a) Muscle contraction follows a series of protein interactions depending upon cellular $[Ca^{2+}]$ (system input). Ca^{2+} activates regulatory units [1 RU=1 troponin (yellow) and 1 tropomyosin (green)], which leads to myosin cross-bridges (red) binding actin (blue) to generate force (contractile output, monitored as force, position, and velocity). Cooperative protein interactions between RUs (RU-RU cooperativity) lead to greater activation in the regulatory protein network, which amplifies contraction. For example, RU_i (activated in b_i) increases the activation probability at neighboring units RU_{i-1} and RU_{i+1} (dashed arrows; b_i), enhancing neighboring activation (b_{ii}).

protein network, reducing coupled protein interactions (i.e., destroying multiple forms of cooperativity) becomes difficult without losing physiological function (because contraction no longer occurs). Incorporating these experiments with complementary, computational simulations controlling cooperative protein interactions and spatial variables of the network will characterize molecular mechanisms producing force in muscle at a level of detail not previously possible.

1) Modeling and Implementation: Computation models provide a platform to interpret and predict coupled molecular events responsible for force production in muscle. These spatially-explicit computational models follow structural, kinetic, and mechanical analyses described more fully by Tanner et al. [29]. Briefly, this computational platform is similar to a finite-element model where a spatial network of regulatory proteins (RUs) interact with a spatial network of motor protein (XBs). Kinetic interactions within the system, driven with Monte Carlo algorithms, lead to force production for a given $[Ca^{2+}]$. The spatially-explicit nature of the model allows us to simulate cooperative interactions between proteins by altering activation kinetics (Figure 13) depending upon kinetic state of neighboring RUs (activated versus inactivated) and XBs (bound versus unbound).

Simulations largely focus on modulating the force-pCa (pCa= $-log_{10}[Ca^{2+}]$) relationship,[4]

$$F = \frac{F_{max}}{1 + 10^{n_H(pCa-pCa_{50})}}, \qquad (8)$$

[4]F_{max} represents the maximal force value. pCa_{50} describes the $[Ca^{2+}]$ that generates half-maximal force (the midpoint), representing Ca^{2+} sensitivity in the system. n_H is the slope at pCa_{50} and is a measure of the cooperativity.
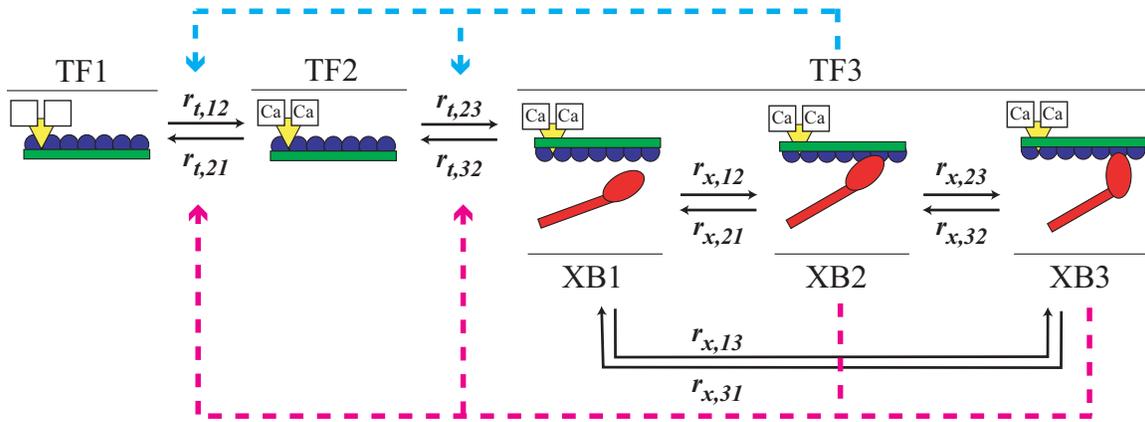
3305

Fig. 13. Simulating RU-RU and XB-RU cooperativity. Coupled kinetics of the regulatory and motor protein networks are described with a three-state $[Ca^{2+}]$-sensitive thin-filament regulatory cycle coupled to a three-state cross-bridge cycle (adapted from [29]). Cooperative interactions from activated RUs (TF3) and/or bound cross-bridges (XB2 or XB3) may influence state transition rates associated with Ca^{2+} binding ($r_{t,12}$, $r_{t,21}$; which have a second order Ca^{2+} dependence) or tropomyosin movement ($r_{t,23}$, $r_{t,32}$). RU activation affecting neighboring RU activation (Figure 12; RU-RU cooperativity) is modeled as a cooperative pathway influencing thin filament kinetics (dashed, cyan arrows). Similarly, the magenta arrows illustrate kinetic influences of XB-RU cooperativity.

to examine how multiple, cooperative protein interactions influence force production. The spatial characteristics of these cooperative mechanisms are precisely controlled in the model, which differs from the experiments. Simulations varied network structure two ways: increasing the RU activation distance in the network (Tm_{span} [29]) and decreasing the number of RUs capable of binding Ca^{2+}.

C. Results

Simulations show that RU-RU or XB-RU cooperativity alone increases Ca^{2+} sensitivity (pCa$_5$0) and slope (n_H), however, the greatest increases in cooperativity come from a synergistic effect combining RU-RU and XB-RU simultaneously (Figure 14). A Tm_{span}=37 nm in a completely functional RU network with cooperative protein interactions (thin dashed line) significantly increased Ca^{2+} sensitivity and slope compared to a similar network with no cooperative interactions (solid line). Between these two conditions, however, the cooperative kinetics slightly compromised maximal force production (at pCa=4.0). Augmenting Tm_{span} to 56 nm, as suggested by skeletal muscle measurements [27], amplifies the maximal predicted force as well as increases cooperativity (greater slope; dotted line). Experiments reducing the number of functional RUs to 20% severely limited force generation ($\diamond$ versus $\circ$). Simulating this structural change produced similar results if the network included cooperative protein interactions (thin dashed versus thick dashed line).

Simulations varying RU activation length (Tm_{span}) and functional RU density show that network architecture (spatial coupling) is an important mechanism affecting cooperative force generation in muscle. While the specific molecular interactions responsible for greatly amplifying force with a small change in $[Ca^{2+}]$ are still being investigated, these results indicate that multiple forms of cooperativity are coupled to control physiological muscle behavior.
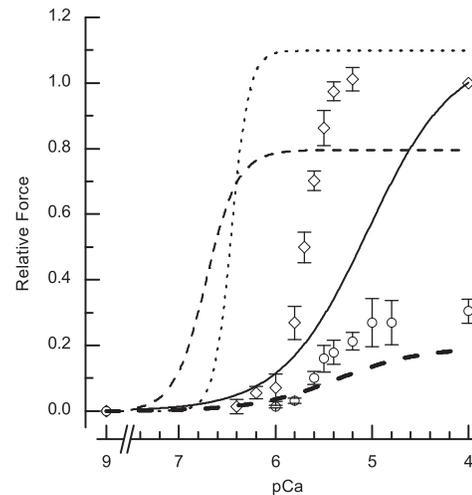

Fig. 14. Experimental measurements of steady-state force versus pCa from rabbit skeletal muscle fibers are shown for a 100% functional regulatory network ($\diamond$) and after exchanging mutant troponin proteins that left only 20% of the RUs functional ($\circ$). Simulating cooperative interactions in a 100% functional regulatory network with Tm_{span}= 37 (thin dashed line) or Tm_{span}=56 nm (dotted line) significantly shifted the force-pCa relationship, compared to a network without cooperative interactions (solid line). Simulations results from cooperative interactions in a network with 20% functional RUs (thick dotted line) reduced force, Ca^{2+} sensitivity, and cooperativity, similar to the experimental results.

D. Conclusions

These computational simulations help design experiments, predict results, and coordinate experimental and theoretical efforts to increase understanding about muscle at a level not previously possible. Changing the activation kinetics or structure of the protein network influences the ability of cooperative protein interactions to amplify force with a small change in $[Ca^{2+}]$. Therefore, controlling these parameters in a patient suffering from a skeletal or cardiac muscle disorder can lead to improved contraction. Potential application

involve genetic, stem cell, or cell transplantation therapies to alter protein kinetics or cellular content in the muscle. Furthermore, similar control targets applied to muscle cells grown *in vitro* could lead to biologically-driven actuators or control devices. Improving models of this spatially-complex system will further determine physiologically relevant control parameters at the molecular or cellular level, where measuring coupled system behaviors independently is difficult.

REFERENCES

[1] Abbondanzieri, E., Greenleaf, W., Shaevitz, J., Landick, R., and Block, S. (2005). Direct observation of base-pair stepping by RNA polymerase. *Nature*, 438(24):460–465.

[2] Akeson, M., Branton, D., Kasianowicz, J., Brandin, E., and Deamer, D. (1999). Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophysical Journal*, 77:3227–33.

[3] Bates, M., Burns, M., and Meller, A. (2003). Dynamics of DNA molecules in a membrane channel probed by active control techniques. *Biophysical Journal*, 84:2366–2372.

[4] Benner, S., Chen, R. J., Wilson, N. A., Abu-Shumays, R., Hurt, N., Lieberman, K. R., Deamer, D. W., Dunbar, W. B., and Akeson, M. (2007). Sequence specific detection of DNA polymerase binding using a nanopore-based state machine. Submitted to *Nature Nanotechnology*.

[5] Bustamante, C., Bryant, Z., and Smith, S. B. (2003). Ten years of tension: single-molecule DNA mechanics. *Nature*, 421:423–427.

[6] Deamer, D. and Branton, D. (2002). Characterization of nucleic acids by nanopore analysis. *Acc. Chem. Res.*, 35:817–825.

[7] Dekker, C. (2007). Solid-state nanopores. *Nature Nanotechnology*, 2:209–215.

[8] Mathé, J., Visram, H., Viasnoff, V., Rabin, Y., and Meller, A. (2004). Nanopore unzipping of individual DNA hairpin molecules. *Biophys. J.*, 87:3205-12.

[9] Rhee, M. and Burns, M. (2006). Nanopore sequencing technology: research trends and applications. *Trends in Biotechnology*, 24:580–6.

[10] Vercoutere, W., Winters-Hilt, S., Olsen, H., Deamer, D., Haussler, D., and Akeson, M. (2001). Rapid discrimination among individual DNA hairpin molecules at single-nucleotide resolution using an ion channel. *Nat Biotechnol*, 19(3):248–52.

[11] Wilson, N. A., Abu-Shumays, R., Benner, S., and Dunbar, W. B. (2007). Rapid finite state machine control of individual DNA molecules in a nanopore. Submitted to *Int. Conf. on Biomedical Electronics and Devices (BIODEVICES)*.

[12] Seelig, G., Soloveichik, D., Zhang, D.Y., and Winfree, E. (2003). Enzyme-free nucleic acid logic circuits. *Science*, 314:1585–8.

[13] Flamm, C., Fontana, W., Hofacker, I.L., and Schuster, P. (2000). RNA folding at elementary step resolution. *RNA*, 6:325–38.

[14] Zuker, M., and Sankoff, D. (1984). RNA secondary structures and their prediction. *Bull. Math. Biology*, 46:591–621.

[15] SantaLucia, Jr., J. (1999). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*, 95:2053–68.

[16] Dirks, R.M., Bois, J.S., Schaeffer, J.M., Winfree, E., and Pierce, N.A. (2007). Thermodynamic analysis of interacting nucleic acid strands. *SIAM Review*, 49:65–88.

[17] Bishop, J.; Klavins E. (2007). *Nanoletters*, in press. DOI: 10.1021/nl070752s.

[18] Chen, Y.; Wang, M.; Mao, C. (2004). *Angew. Chem. Int. Ed.*, 43, 3554.

[19] Koumura, N.; Zijlstra, R.W.; van Delden, R.A.; Harada, N.; Feringa, B.L. (1999). *Nature*, 401, 152.

[20] Fennimore, A.M.; Yuzvinsky, T.D.; Han, W.Q.; Fuhrer, M.S.; Cumings, J. (2003). *Nature*, 424, 408.

[21] Regan, B.C.; Aloni, S.; Jensen, K.; Zettl, A. (2005). *Appl. Phy. Lett.*, 86, 123119.

[22] Balzani, V.; Clemente-Leon, M.; Credi, A.; Ferrer, B.; Venturi, M.; Flood, A.H.; Stoddart, J.F. (2006). *Proc. Natl. Acad. Sci.*, 103, 1178.

[23] Turberfield, A.J.; Mitchell, J.C.; Yurke, B.; Mills, Jr., A. P.; Blakey, M. I.; Simmel, F. C. (2003). *Phys. Rev. Lett.*, 90, 118102.

[24] Yin, P.; Yan, H.; Daniell, X.G.; Turberfield, A.J.; Reif, J.H. (2004). *Angew. Chem. Int. Ed.*, 43, 4906.

[25] Santoro, S.W.; Joyce, G.F. (1997). *Proc. Natl. Acad. Sci.*, 94, 4262.

[26] Miller, H.I.; Gill, G.N.; Riggs, A.D. (1973). *J. Biol. Chem.*, 248, 2621.

[27] Regnier, M., Rivera, A.J., Wang, C.K., Bates, M.A., Chase, P.B. and Gordon, A.M. (2002). Thin filament near-neighbour regulatory unit interactions affect rabbit skeletal muscle steady-state force-Ca(2+) relations. *J Physiol*, 540(Pt 2):485–497.

[28] Daniel, T.L., Trimble, A.C. and Chase, P.B. (1998). Compliant realignment of binding sites in muscle: transient behavior and mechanical tuning. *Biophys J*, 74(4):1611–1621.

[29] Tanner, B.C.W., Daniel, T.L., and Regnier, M. (2007). Sarcomere Lattice Geometry Influences Cooperative Myosin Binding in Muscle. *PLoS Comput Biol*, 3(7):e115.

[30] Gordon, A.M., Homsher, E., and Regnier, M. (2000). Regulation of contraction in striated muscle. *Physiol Rev* 80(2):853–924.

[31] Gillis, T.E., Martyn, D.A., Rivera, A.J., and Regnier, M. (2007). Investigation of thin filament near-neighbour regulatory unit interactions during force development in skinned cardiac and skeletal muscle. *J Physiol* 580(2):561–576.