## Data Analysis

The data analysis and the step average performed through the paper rely on the synchronizing procedure. In this supplementary material we describe the algorithms used to find and align the steps and we test some of their features against data obtained from stochastic simulations. Moreover, we test the model presented in the *Discussion* of the main article and show that the algorithm predicts the ATP dependent 5 nm displacement observed in the experiments.

**Finding Algorithm.** The steps were extracted after the original data had been separated in many sequence of "staircases" that look like Fig. 2A in the main article. Steps were detected using a "sliding variance" filter (1), the maxima of which correspond to the step positions. Then we selected the steps, whose variance exceeded a manually fixed threshold. This algorithm is not completely unsupervised: for example, multiple steps, backward steps, or incomplete steps were manually discarded from the statistics.

**Synchronization Algorithm.** Once the steps were identified, they were cut within a window of $\pm 10$ ms around the main displacement. In order to synchronize in such a way that the main displacement occurs at the same time $t_0$ for all the steps, we proceeded as follows. The mean values of the two plateaus, before and after the step, were used to determine respectively the initial and the final position of the motor. A region of 3.2 ms around the step was excluded to avoid unexpected artifacts. Then, each step was fit with the response function of a bead subjected to a step-like force (see Eq. **1** in the main article) of amplitude $A$:

$$\chi(t) = x_0 + A\left[1 - \exp\left(-\frac{t - t_0}{\tau}\right)\right], \qquad [1]$$

where the only fitting parameter is $t_0$. The response time $\tau$ is different for each curve since we are analyzing an ensemble of data at different forces, but for simplicity we consider the reasonable value $\tau = 100\ \mu$s, which is compatible with the data extracted from power spectrum analysis. Changes in the value of $\tau$ in the range of hundred microseconds does not influence too much the efficiency of the algorithm. Note: steps shorter than 20 nm or larger than 45 nm have been discarded from the statistics.

## Test of the Algorithms

The synchronization algorithm could introduce artifacts in the data analysis. In particular, the fitting is rather sensitive to the steepness of the curve and can easily confuse a large fluctuation with the real step. In order to estimate the error introduced by the synchronization, we show both by analytical calculations and by numerical simulations that the misalignment made by the algorithm for different models is comparable with the response time of the bead (so below the effective resolution of our measurement).

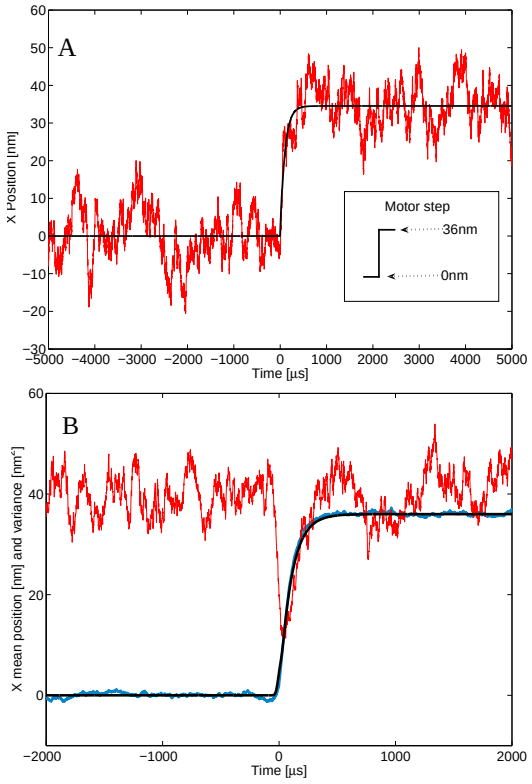


**Fig. 1.** One-Phase Model: the motor steps directly from 0 nm to 36 nm. (*A*) Simulated step (red) with $\kappa = 0.1$ pN/nm and $\xi = 10$ pN·$\mu$s/nm (i.e. $\tau = 100\ \mu$s); in black the fit with the response function $\chi$. Inset: scheme of the motion of the motor. (*B*) Average and variance of the simulated steps after the alignment procedure and selection of the steps longer than 20 nm and shorter than 45 nm. Superimposed in solid black is the average step computed analytically (Eq. **19**). As described in the text the variance deviates from a constant value only around the alignment point on a time scale comparable to the bead responses time ($\sim \tau$). Moreover, in the mean step there is no evidence of a prestep: the algorithm does not introduce any artifact in the one-phase model.

**Numerical Simulations.** In order to test the validity of the algorithm on physical parameters we performed stochastic simulations. We consider the motion of the bead always subjected to thermal noise. The position of the bead $x(t)$ is derived from the Langevin equation assuming the bead to be bound to the motor (in position $x_m(t)$)) through an elastic spring of constant $\kappa$, hence confined in the harmonic potential $U = \kappa(x - x_{\rm m})^2$.

$$\xi\dot{x}(t) + \kappa(x - x_{\rm m}) + \eta(t) = 0\,, \qquad [2]$$

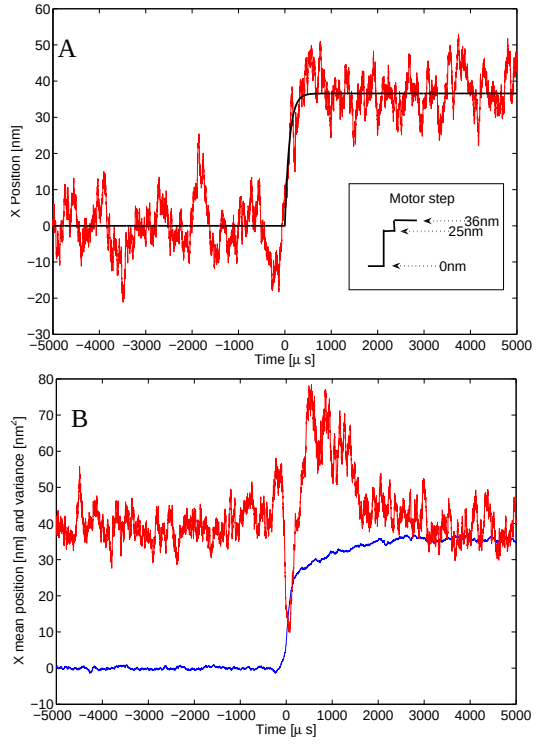


**Fig. 2.** Two-phase model: after a first deterministic step to 25 nm the motor moves stochastically to 36 nm. A) Simulated step (red) with $\kappa = 0.1$ pN/nm and $\xi = 10$ pN·$\mu$s/nm (i.e. $\tau = 100$ $\mu$s); in black the fit with the response function $\chi$. *Inset*: scheme of the motion of the motor. (*B*) Average and variance of the simulated steps after the alignment procedure and selection of the steps longer than 20 nm and shorter than 45 nm. As in the previous model, the variance deviates from a constant value only around the alignment point on a time scale comparable to the bead responses time ($\sim \tau$). The fact that the mean step does not show any prestep (seen also in other data not shown) allows us to think that the algorithm is not introducing any artifact in a two-state model.

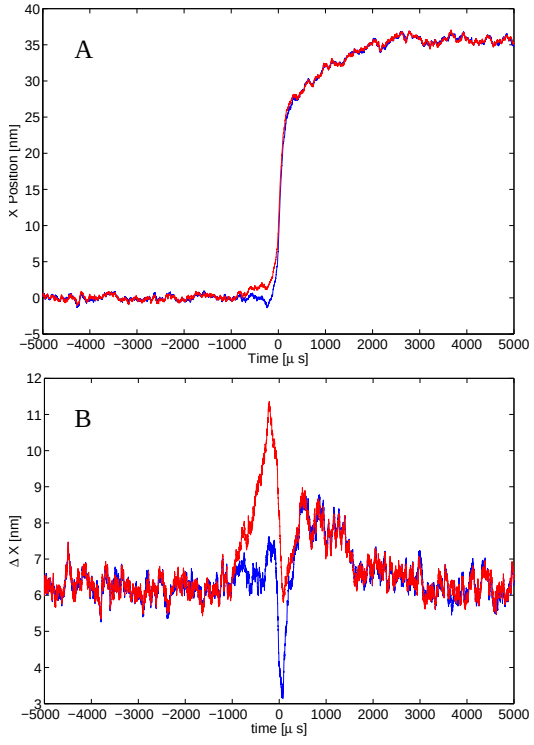


**Fig. 3.** Two-phase model: simulated vs artificially shifted steps. We compared simulated data of the double phase model aligned with the synchronization algorithm (blue) and artificially shifted (red). (*A*) Average step: the shifted data show clearly the slope preceding the main displacement, as found in the experimental data. (*B*) Variance of the step: the shifted step gives an unrealistic variance that increases much stronger than experimental one in the phase preceding the displacement from 5 nm to 25 nm.

where $\xi$ is the drag coefficient of a 200-nm bead in the water. We integrated numerically such an equation generating at each time step a random number properly normalized (corresponding to the random force $\eta$) and using a simple Euler algorithm. The position of the motor $x_{\rm m}$ has been determined before each step and has been chosen in two different ways: the one- and two-phase models.

**One-Phase Model.** In the first simplest model we consider the motor to move forward 36 nm deterministically at the time $t_0$:

$$x_m(t) = \begin{cases} 0 \text{ nm if } t < t_0 \\ 36 \text{ nm if } t \geq t_0 \end{cases} \qquad \textbf{[3]}$$

*Inset SI* in Fig. 1*A* shows the shape of the motor displacement. In the same figure we show one simulated step (red line) and the relative fit (black line), for $\kappa = 0.1$ pN/nm and a response time $\tau = 100$ $\mu$s. As can be seen, the steps (yet very noisy) do resemble the experimental ones. We note that the fluctuations are strictly gaussian, which is not always the case in the experimental data. The average step is shown in *SI* Fig. 1*B* and, in contrast to the experimental results, does not evidence any kink or substep (compare to Figs. 3 and 4 in the main article). The superimposed black line corresponds to the average step computed analytically in the last paragraph of this document (see Eq. **19**). Note that the variance does not increase after the step and deviates from a constant value only around the fitting value $t_0$ in a time window of $\sim 2\tau$: the error committed by the algorithm is comparable to the time of the response of the bead; we also prove this point by analytical calculations (see below). We conclude that the algorithm does not introduce any artifact if no substeps are present.

**Two-Phase Model.** In the second model, inspired by experimental evidences (2-5), we supposed that the motor moves forward in two stages: first it steps over 25 nm in a deterministic way, then it completes the motion to 36 nm. The time delay is statistically distributed and follows a Poisson law, with a time scale of 1 ms:

$$x_{\rm m}(t) = \begin{cases} 0 \text{ nm if } t < t_0 \\ 25 \text{ nm if } t_0 \leq t < t_1 \\ 36 \text{ nm if } t \geq t_1 \end{cases} \qquad \textbf{[4]}$$

*Inset* in Fig. 2*A* shows the shape of the motor trajectory. In the same figure we show an example of a simulated step (red line) and the relative fit (black line). Although each step resembles the previous ones (one-phase model), the average step clearly shows the exponential relaxation from 25 to 36 nm (see Fig. 1*B*) and well reproduces the experiments. In contrast to the experimental results, the average does not evidence any substep before the main jump (compare to Figs. 3 and 4 in the main article). The variance increases after the step and decays on the time scale of $\tau_1$ (as expected form a convolution of a step function with an exponential distribution). Now, let us suppose that the algorithm erroneously fit the beginning of the main displacement in $t_1$. In Fig. 3 we intentionally shifted 10% of the steps and

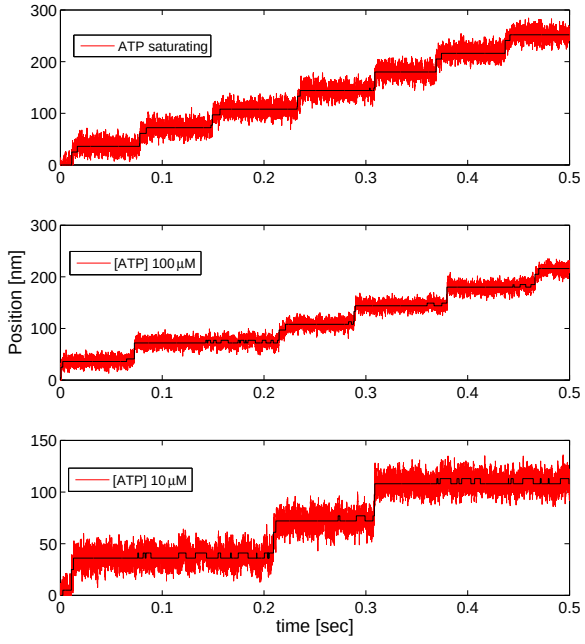


**Fig. 4.** Three-phase model: the motor switches stochastically and reversibly from state A (at 0 nm) to state B (at 5 nm); only from this state, upon the arrival of an ATP molecule the motor can switch to state C (25 nm) and finally to the next plateau at 36 nm. Simulated response to the bead (red) are compared to the simulated motion of the motor (black solid line) for different ATP concentration (2 mM, 100 $\mu$M, and 10 $\mu$M, respectively).

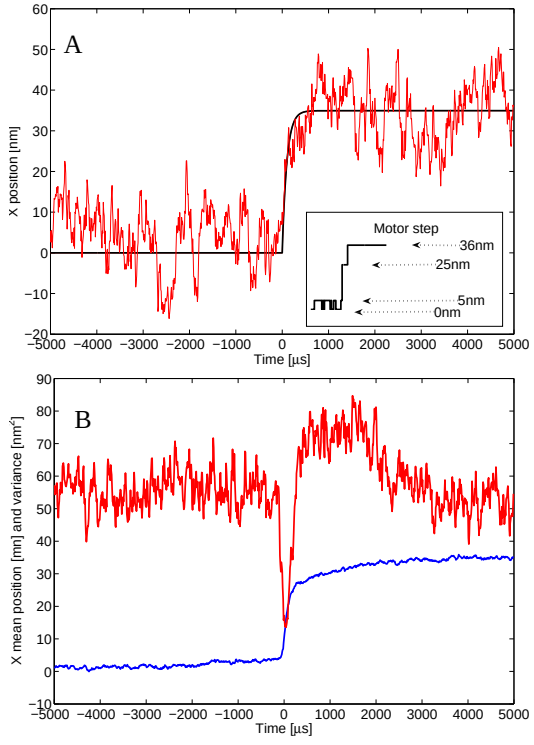


**Fig. 5.** Three-phase model, with steps at 0 nm, 5 nm and 25 nm. (*A*) Simulated step (red) with a 7nm peak-to-peak white noise and response time of the bead of $\tau = 100$ $\mu$s; in black the fit with the response function $\chi$. (*B*) Average and variance of the simulated steps after the alignment procedure and selection of the steps longer than 20 nm and shorter than 45 nm. In this case the average step resembles the one obtained experimentally. The variance increases after the main displacement (26 nm to 36 nm step) but, in contrast with the two-phase model, does not change significatively before $t_0$.

introduce therefore a systematic error in the alignment. An artificial substep of about 5 nm indeed appears in the average step. Although the average step looks like the experimental data, the variance remarkably increases before the main displacement. This was not the case in the experimental data, where the variance increased only *after* the main step, due to the diffusive search (step from 25 nm to 36 nm). Eventually, the variance allows one to discriminate between a real prestep (present in every single realization) and a statistical effect (emerging from an average of misaligned data). We conclude that our algorithm is rather accurate in finding the beginning of the step even though the fit function $\chi$ is oversimplified and does not fit perfectly the unknown shape of the steps.

**Three-Phase Model.** In our third model we consider a motion of the motor that gives a realistic average and variance for the steps and that is compatible with the enzymatic data (6) As explained in the main article we suppose the first phase leading from 0 nm to 5 nm to be the average of sharp steps of different length. We suppose that, after the ADP release, the motor switches *stochastically* and *reversibly* between the two states A (0 nm) and B (5 nm). The main step toward state C will occur only if two conditions are simultaneously verified: (*i*) the motor is in state B *and* (*ii*) the ATP is bound. Transition from C to A' = A + 36 nm occurs in a diffusive manner as explained in the previous model. In our simulation the minimum dwell time has been fixed to 60 ms, according to the enzymatic data. During this time the motor waits in the state A. Since the energy barrier between state A and B is unknown, we have chosen the simplest set of parameters compatible with experiments and kinetics rates. The switching rate $\omega_{\mathrm{A}\rightleftharpoons\mathrm{B}}$ has been fixed to 200 $\mathrm{s}^{-1}$ in both directions, and the rate $\omega_{\mathrm{B}\to\mathrm{C}}$ varies from 10 to 1000 $\mathrm{s}^{-1}$. When the time of arrival of ATP is shorter than the time scale of the switch, almost no prestep is visible in the average step. On the contrary, at low ATP concentration, the time scales mix, since the transition $\mathrm{A}\to\mathrm{C}$ emerges from the two processes $\mathrm{A}\rightleftharpoons\mathrm{B}$ and $\mathrm{B}\to\mathrm{C}$. Eventually (but this is beyond the scope of this paper) this simple model accounts also for the force velocity relation (P.P. and G.C., unpublished work).

From Fig. 5 one can observe that the average step fits the experiments much better than the previous models. Moreover, the variance does not increase before phase 2. The slope from 0 nm to 5 nm, occurring on the millisecond time scale, is therefore an average of uniformly distributed sharp steps from 0 nm to 5 nm. Figure 6 highlights how the length of phase 1 scales with the ATP concentration. In particular, we notice that this time scale saturates when the arrival of an ATP molecule is too rare ($\omega_{\mathrm{A}\rightleftharpoons\mathrm{B}} \gg \omega_{\mathrm{B}\to\mathrm{C}}$).

## Analytical Calculations of the Synchronization Error

In the hypothesis that the motor moves in a step-like manner like in the one-phase model, the synchronization uncertainty can be analytically computed.

The reasoning line is the following. The motor position is de-

scribed by Eq. (**3**) with amplitude $A = 36$ nm:

$$x_{\rm m}(t - t_0) = A\theta(t - t_0)\,. \qquad \textbf{[5]}$$

We can write the solution of the Langevin equation (Eq. **2**) as the sum of a deterministic part $\chi(t - t_0)$ and a stochastic one due to thermal fluctuation $x_{\rm th}(t)$:

$$x(t) = \chi(t - t_0) + x_{\rm th}(t) \qquad \textbf{[6]}$$

with

$$\chi(t - t_0) = A\theta(t - t_0)\left(1 - e^{-(t-t_0)/\tau}\right), \qquad \textbf{[7]}$$

$$x_{\rm th}(t) = \int_{-\infty}^{t} \eta(t')e^{-\frac{t-t'}{\tau}}dt'\,. \qquad \textbf{[8]}$$

The response time of the bead is $\tau = \xi/\kappa$. The noise $\eta$ has been chosen to be white, with zero average and variance given by the fluctuation-dissipation theorem, i.e. its probability distribution function is

$$P(\eta) = \frac{1}{\sqrt{4\pi k_{\rm B}T\xi}}\exp\left(-\frac{\eta^2}{4k_{\rm B}T\xi}\right), \qquad \textbf{[9]}$$

$\langle\eta(t)\rangle = 0$ and $\langle\eta(t)\eta(t')\rangle = 2k_{\rm B}T\xi\delta(t - t')$. Since the noise averages to zero, we fit the function $x(t)$ with $\chi(t - \tilde{t})$ where $\tilde{t}$ is the fitting parameter. According to the least square method we wish to minimize the function

$$\int_{-\infty}^{\infty}\left(\chi(t - t_0) + x_{\rm th}(t) - \chi(t - \tilde{t})\right)^2 dt \qquad \textbf{[10]}$$

since any step is considered independent in the infinite interval.

The least square requires:

$$\frac{\partial}{\partial\tilde{t}}\int\left(\chi(t - t_0) + x_{\rm th}(t) - \chi(t - \tilde{t})\right)^2 dt = 0\,. \qquad \textbf{[11]}$$

This can be rewritten as sum of three terms:

$$2\int\frac{\partial\chi(t - t_0)}{\partial\tilde{t}}\left(\chi(t - t_0) + x_{\rm th}(t) - \chi(t - \tilde{t})\right)dt = 0\,. \qquad \textbf{[12]}$$

The integration can be performed exactly up to the noisy term and results in the two cases:

$$e^{-(\tilde{t}-t_0)/\tau} = 1 + \frac{2}{A\tau}\int_{\tilde{t}}^{\infty} e^{-(t-\tilde{t})/\tau}x_{\rm th}(t)dt \quad \text{if} \quad \tilde{t} > t_0$$

$$e^{-(t_0-\tilde{t})/\tau} = 1 - \frac{2}{A\tau}\int_{\tilde{t}}^{\infty} e^{-(t-\tilde{t})/\tau}x_{\rm th}(t)dt \quad \text{if} \quad \tilde{t} < t_0$$

**[13]**

We can expand the exponential in the condition that $|\tilde{t} - t_0| \ll \tau$. This (rather crude) approximation provides

$$\delta t \equiv t_0 - \tilde{t} \approx \frac{2}{A}\int_{\tilde{t}}^{\infty} e^{-(t-t')/\tau}x_{\rm th}(t)\,. \qquad \textbf{[14]}$$

Performing the average and inserting the random trajectory described by Eq. **8** one obtains

$$\langle\delta t^2\rangle \approx \frac{4}{A^2}\int_0^{\infty}dt\int_0^{\infty}dt'e^{-2(t+t')/\tau}\int_{-\infty}^{t}dt''\times$$
$$\times\int_{-\infty}^{t'}dt'''\left\langle\eta(t'' - \delta t)\eta(t''' - \delta t)\right\rangle e^{(t''+t''')/\tau}. \qquad \textbf{[15]}$$

Using the condition on the variance of the noise (that provides a $\delta$-function)

$$\langle\delta t^2\rangle \approx \frac{8k_{\rm B}T}{A^2\xi}\int_0^{\infty}dt\int_0^{\infty}dt'e^{-2(t+t')/\tau}\int_{-\infty}^{t}dt''\times$$
$$\times\int_{-\infty}^{t'}dt'''\delta(t'' - t''')e^{(t''+t''')/\tau}. \qquad \textbf{[16]}$$

Finally, integrating on the four variables we obtain:

$$\langle\delta t^2\rangle \approx \frac{2k_{\rm B}T}{A^2\kappa}\tau^2 \qquad \textbf{[17]}$$

For the parameters set in the simulations, which roughly correspond to the experimental ones (i.e. $\kappa \sim 0.1$ pN/nm and $A = 36$ nm) the uncertainty of the algorithm is $\delta t \sim \tau/4$.

The average step can now be analytically computed as a convolution of the bead response function $\chi(t - t_0)$ (see Eq. **7**) with a gaussian distribution of width $\sigma = \sqrt{\langle\delta t^2\rangle}$:

$$\langle x(t)\rangle = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{t}\chi(t - t_0 - \delta t)e^{-\frac{\delta t^2}{2\sigma^2}}\,d\delta t. \qquad \textbf{[18]}$$

Given $t_0 = 0$, this leads to an average step

$$\langle x(t)\rangle = F\left(\frac{t}{\sigma}\right) - F\left(\frac{t}{\sigma} - \frac{\sigma}{\tau}\right)\exp\left(-\frac{t}{\tau} + \frac{\sigma^2}{2\tau^2}\right), \qquad \textbf{[19]}$$

where $F(t)$ is the standard normal cumulative distribution function, defined as $F(t) \equiv \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}e^{-x^2/2}\,dx$.

## Conclusions

In conclusion, we proved both analytically and numerically that the accuracy of the fitting algorithm is in the order of magnitude of $\tau$: hence any phenomenon occurring on larger time scale is not a statistic artifact introduced by the algorithm itself. A simple three-phase model, compatible with kinetics rates, energetic scales, experimental average steps, and variance, accounts for the prestep observed in our experiments.

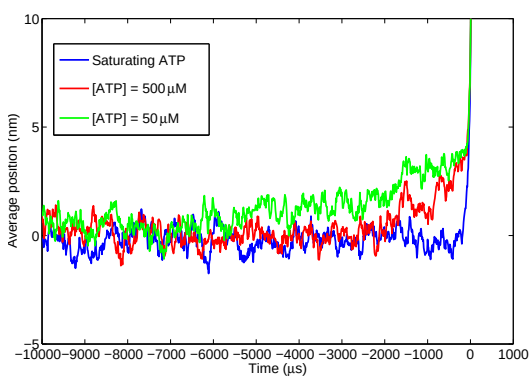


**Fig. 6.** Three-phase model: prestep in the average step for different ATP concentrations. According to our model, the ATP arrival time determines the length of the prestep. On average this gives linear profiles, whose slopes depend on the ATP concentration.

1. Smith DA (1998) *Phil Trans R Soc Lond B* 353:1969-1981.
2. Veigel C, Wang F, Bartoo ML, Sellers JR, Molloy JE (2002) *Nat Cell Biol* 4:59-65.
3. Spudich JA, Rock RS (2002) *Nat Cell Biol*, 4:E8-E10.
4. Uemura S, Higuchi H, Olivares AO, De La Cruz EM, Ishiwata S (2004) *Nat Struct Mol Biol*, 11:877-883.
5. Walker ML, Burgess SA, Sellers JR, Wang F, Hammer JA, Trinick J, Knight PJ (2000)*Nature*, 405:804-807.
6. De La Cruz EM, Wells AL, Rosenfeld SS, Ostap EM, Sweeney HL (1999) *Proc Natl Acad Sci USA* 96:13726-13731.