Supplementary Methods

Calibration

Internal Distance Standard and Systematic Errors. A semi-passive mode, see Supplementary Figure 1, provides a means by which the contour length calculation can be cross-checked. In an “active” mode measurement the separation of the system is controlled dynamically to maintain a constant optically applied force\(^1\). In such a measurement, the packaged contour length is determined directly from this change in separation and the elastic behavior of the tether at the known force. In a “passive” mode measurement, in which the trap separation is constant, the contour length is calculated from the displacement change of each bead, the force change, and the change in the absolute extension of the molecule\(^2\). Since each packaging segment will pass through the same optically applied tensions as the previous segment, effectively creating a two point active mode measurement between segments, semi-passive mode measurements allow both type of calculations to be performed. This provides a means by which the accuracy of the contour length calculation can be determined.

Explicitly, the observed change in contour length \(\Delta L_{\text{obs}}\) between the points at which the applied tension reaches a value \(F_0\) in two adjacent passive mode segments is determined by adding three contour length changes \(\Delta L_1\), the observed contour length change from the force set point to the end of the packaging segment, \(\Delta L_2\), the observed contour length from the beginning of the next segment to the contour length at which the tension reaches the same force value, and \(\Delta L_3\), the contour length packaged during the reset time, \(\Delta t\), estimated from the observed velocity of the initial segment, i.e. \(\Delta L_3 = v\Delta t\) (Supplementary Figure 1.) This sum, reflecting
contour lengths calculated in the passive mode fashion, should be equal to the contour length change \( \Delta L_{\text{pred}} \) predicted from the change in trap separation \( \Delta d \) between segments, which is calculated directly via \( \Delta L_{\text{pred}} = \Delta d / \alpha(F_0) \). \( \alpha(F_0) \) is the ratio of the extension to the contour length at the force \( F_0 \) calculated from the extensible-worm-like-chain model with the parameters above. The relative error between these values is calculated by

\[
\delta = \frac{\Delta L_{\text{obs}} - \Delta L_{\text{pred}}}{\Delta L_{\text{pred}}}.
\]

Because the motion of the steerable mirror is calibrated to <1% and stable to <1 Å\(^3\) and errors in the force result in minor errors in the parameter \( \alpha(F_0) \) for forces above ~6 pN (at 6 pN a 10% error in the force results in <1% error in the value \( \alpha(F_0) \)), the predicted contour length can be determined with very high precision and accuracy, ~1%. In practice, stochastic errors in \( \Delta L_{\text{obs}} \) are much larger than 1%; however, since the measured value of \( \delta \) for each semi-passive mode segment is largely independent of the contour length of DNA packaged (data not shown), the values of \( \delta \) from all packaging segments under a given set of conditions are averaged to produce an estimate of systematic errors.

Supplementary Figure 2 shows the systematic error calculated from the average \( \delta \) for all passive mode segments. For all low force values, except for one outlier, the systematic errors appear to be independent of [ATP] and laser wavelength, suggesting that these errors most likely originate in the assumption of the drag coefficient for thermal spectra calibration. This is supported by the large difference in the systematic error when the packaging buffer is composed of 80% D\(_2\)O, where a large change in the viscosity of the solvent occurs. Because the stochastic error in these values are small and because \( \Delta L_{\text{pred}} \) can be estimated with high precision and accuracy, all data reported in this paper, except for raw stepping data, have been corrected for these small systematic errors.

**Selection of Low Noise Data.** It has been previously noted that the noise characteristics of individual bead-DNA-bead systems vary for reasons that are not completely understood\(^3\)-\(^6\). We see this effect in our own data as well. To circumvent this, it has become common practice to select data based on low noise characteristics, with meaningful results being inferred from fractions of the data as low as 10%\(^6\). Such practices have the possibility of ignoring the global behavior of a system while focusing on behaviors that may only exist in a small population.
To avoid this issue, we start by investigating the average PWD for all data collected under given conditions, including data that have obviously larger noise characteristics. If periodicity is observed in this global PWD (See Supplementary Figure 3), we then sort individual semi-passive mode segments by the strength of the global periodicity, i.e., the clarity of the stepping behavior. This is measured by the integrated area in a specific region of the Fourier transform of the PWD, corresponding to 9 to 11 bp in the low force data and 2.3 to 2.7 bp for the high force data. The PWD that we report in the main text is the average of the PWD from the top 50% of semi-passive mode segments selected in this fashion. These selected traces are used for all subsequent dwell time analysis. Supplementary Figure 2 lists the number of packaging complexes, the total number of passive mode segments, and the total contour length packaged by the top 50% of packaging complexes selected in this fashion and used in all analysis presented. While it is still possible that we are missing the dynamics of a subpopulation, Supplementary Figure 3 clearly shows that the dynamics we do select and analyze represent the dominant behavior of the packaging motor of φ29.
Supplementary Discussion

Correlation Analysis. Supplementary Figure 4a shows that the 10-bp bursts observed at low external force are occasionally interrupted by micro-pauses, strongly suggesting that the 10-bp packaging increments are composed of smaller steps. While PWD do not detect these events because peaks in a PWD are weighted by the product of the average time spent in each step, the step finding algorithm based on the t-test does. Supplementary Figure 4b and c show the two dimensional probability distribution of observing a specific step size with a specific dwell time determined from the step finding algorithm. The distribution is clearly bimodal. The first mode has an average step size of ~10 bp and a dwell time distribution that is peaked, region A in Supplementary Figure 4c. These events correspond to the 10-bp bursts analyzed in Figures 1 and 2. The second mode corresponds to a second population of steps, with average size ~5 bp. The dwell time distribution for these events appears to be a weighted sum of a peaked dwell time distribution, consistent with that observed for the 10-bp bursts, and an exponential distribution. This is consistent with an interrupted burst model since an observed micro-pause would separate the 10-bp step into smaller steps, the first of which would have a dwell time drawn from the 10-bp burst dwell time distribution while the second would be drawn from a distinct distribution.

As further support for a model in which the micro-pauses interrupt the 10-bp bursts, we performed a correlation analysis in which we calculate the step size distributions conditional on observing a following step of a certain type. We divide the smaller steps into two regions, a region of small steps with long dwell times, region B, and a region of small steps with short dwell times, region C (see Supplementary Figure 4c.) The division was selected at the point of minimum overlap between the two distributions to minimize the probability of misidentifying each step. We then calculate the step size distributions conditional on observing a following step from each region, shown in Supplementary Figure 4d. The dramatic increase in the probability of the smaller steps that precede events in region C—short dwells preceding a small step—as compared to the probability of a ~10-bp step strongly suggests that these smaller steps occur in pairs, the first small step with a long dwell time and the second small step with a short dwell time. The conditional distributions for steps preceding events in regions A and B are very similar to the non-conditional distribution. This is also consistent with a model in which the micro-pauses represent an interrupted burst since the probability of the preceding 10-bp burst
being interrupted is independent of whether or not the current 10-bp burst is interrupted, i.e. if the step is in region A or B. This analysis rules out the alternative interpretation of our data—that ϕ29 has a force dependent step size that varies from 10 bp at low force to 2.5 bp at high force—since in this case one would not expect to see that the observed small steps are correlated in time. This is further supported by the lack of a force dependence of the 2.5-bp steps, seen in Figure 3b.

The fact that the smaller step distribution has an average size of ~5 bp is most likely an artifact due to the greater ease of detecting an interrupted burst when the 3rd step has a longer dwell than when either the 2nd or 4th steps have the long dwell time. In the first case one needs the resolution to observe a 5-bp transition while in the later case a 2.5-bp resolution is needed, a resolution not attained at the low tensions and high bandwidths of these experiments.

**D2O Does Not Change the Burst Size.** In the course of the initial high force experiments, it was discovered that the increased laser heating of the sample due to the required high laser powers resulted in a temperature dependent increase in the packaging velocity (data not shown). This increase compensated for the expected force dependent decrease in the velocity, prohibiting the frequent observation of the steps that compose the burst. This problem was circumvented by packaging in a buffer composed of 80% D2O (made by diluting 10x packaging buffer into 90% D2O; Sigma-Aldrich, St. Louis, MO) which absorbs significantly less of the 1064-nm trapping light than H2O8, 9.

To demonstrate that the mechanics of packaging are not changed by D2O, we repeated the low force experiments in a 250 μM [ATP], 80% D2O packaging buffer, the same buffer used for high force experiments. Supplementary Figure 5 shows that under these conditions, packaging occurs with bursts of average size 9.7 ± 0.2 bp (s.d.), consistent with the 10.0 ± 0.2 bp bursts observed in H2O. The dwell time distribution under these conditions is also peaked with an $n_{\text{min}}$ of 3.4 ± 0.4 (s.e.m.) consistent with the value observed in H2O, 3.8 ± 0.3, but with an average dwell time of 202 ± 6 ms (s.e.m.), ~1.6 times larger than that observed in an H2O buffer, 124 ± 2 ms. Thus, the presence of D2O does not change the burst size or the number of rate-limiting events at saturating [ATP] but does change the average rate of these kinetic transitions. A similar reduction of the ATPase rate in D2O has been observed for other members of the
ASCE family\textsuperscript{10} and is most likely the result of the solvent isotope effect\textsuperscript{11}. It should be noted that because of the possible effect of D\textsubscript{2}O on the reaction kinetics, the dwell times observed in Figure 3c should not be interpreted as the expected dwell times in H\textsubscript{2}O.

**The High Force Dwell Time Distribution is a Weighted Sum of Two Distributions.** The composite nature of the 10-bp bursts in combination with the bi-exponential nature of the dwell time distribution for the 2.5-bp steps at high force suggests that this distribution may be the weighted sum of two distributions, (1) an exponential decay representing the stepping rate of three of every four 2.5-bp steps, and (2) a peaked distribution similar to that observed at low force for one out of every four steps. This interpretation is supported by the fact that the low force dwell time distribution has a similar exponential tail to the slow exponential in the high force distribution. Fitting the tail of the low force distribution in Supplementary Figure 5b (taken in a D\textsubscript{2}O buffer, as with the high force data) yields an exponential rate of 9.4 ± 0.4 s\textsuperscript{-1} (s.d.), consistent with the slow process observed at high force, 8 ± 1 s\textsuperscript{-1} (s.d.). Furthermore, the relative probability of the fast to slow processes at high force, 1.1 ± 0.3 (s.d.), can be corrected for the increased probability of events in the tail of a peaked distribution, yielding a relative fraction, 2.6 ± 0.8 (s.d.), of fast events drawn from an exponential distribution to slow events drawn from the peaked distribution. Thus, the probability of drawing from the slow distribution is 0.27 ± 0.06 (s.d.), consistent with the expected 1 in 4 probability. Supplementary Figure 6b demonstrates that scaling the low force dwell time distribution with this value produces good agreement between the exponential tails of the two distributions. Moreover, when scaled in this fashion the peak of the low force distribution is hidden by the more probable exponential distribution, producing a dwell time distribution at high force that is well described by a bi-exponential decay.

This analysis further suggests that the individual rates of the three fast dwells are not significantly different and that the 22 ± 2 s\textsuperscript{-1} rate of the fast process represents the average rate for these three steps.

**Michaelis-Menten Velocity with Multiple ATP Bindings.** As described in the text, a motor with multiple substrate binding sites will generally exhibit sigmoidicity in its substrate saturation behavior, except under particular circumstances. To illustrate the kinetic conditions in which sigmoidal binding curves are reduced to simple, Michaelis-Menten behavior, we consider a simplified system with two ATP binding sites. As detailed in the text, we describe ATP...
binding by a two-step process in which ATP loosely docks (denoted by the state $T$), then tightly binds (denoted by state $T^*$) to the catalytic site. In the first model kinetic scheme, we consider that binding of each ATP occurs successively in time and in an ordinal sequence in which the first catalytic site always loads before the second:

$$EE \xrightleftharpoons[k_1]{} k_i[T] \xrightleftharpoons[k_2]{} TE \xrightleftharpoons[k_3]{} T^*E \xrightleftharpoons[k_4]{} T^*T \xrightleftharpoons[k_5]{} T^*T^* \xrightleftharpoons[k_6]{} EE,$$

where $E$ denotes the empty or apo state, $T$ the loosely bound ATP state, and $T^*$ the tightly bound ATP state. $k_i$ denote the forward rates while $\tilde{k}_i$ denote the reverse rates, and the docking rates are expressed as pseudo-first rates, i.e. $k_i[T]$. The final transition also corresponds to the generation of step of size $d$. To be completely general, we allow all transitions to be reversible, in particular we allow for the possibility that product molecules can rebind and the motor can take a backwards step.

It is straightforward to calculate the average velocity from this kinetic scheme\cite{7,12}, which takes the form

$$v = \left( -da_0 + da_2[T]^2 \right)/(b_0 + b_1[T] + b_2[T]^2),$$

and contains terms quadratic in the ATP concentration. It is instructive to consider the explicit dependence of two of these coefficients on the individual rate constants

$$a_0 = \tilde{k}_1\tilde{k}_2\tilde{k}_3\tilde{k}_4\tilde{k}_5$$
$$b_0 = \tilde{k}_1\tilde{k}_2\left(k_2\tilde{k}_5 + k_4\tilde{k}_3 + k_3\tilde{k}_4 + \tilde{k}_3\tilde{k}_4 + \tilde{k}_4\tilde{k}_5\right)$$
$$+ \tilde{k}_1\tilde{k}_3\tilde{k}_4\left(k_2 + \tilde{k}_1 + \tilde{k}_2\right)$$

Note that if either $\tilde{k}_1 = 0$ or $\tilde{k}_2 = 0$ and $\tilde{k}_3 = 0$, $\tilde{k}_4 = 0$, or $\tilde{k}_5 = 0$ these coefficients will vanish and this expression will reduce to the simple Michaelis-Menten substrate saturation behavior $v = d a_2[T]/(b_1 + b_2[T])$. This represents a general property of any linear kinetic scheme with multiple ATP binding transitions. If these binding events are connected via reversible transitions the system will display a velocity which depends on powers of the substrate concentration higher
than 1. However, if all of these transitions are separated by irreversible transitions, then the velocity is reduced to a simple Michealis-Menten $[T]$ dependence.

Previous work has shown that ADP and non-hydrolyzable ATP analogs can unbind from the packaging motor$^{13}$, thus we must require that $\bar{k}_1 \neq 0$ and $\bar{k}_3 \neq 0$. And, to explain our data, we must invoke a second binding transition that follows each docking transition, separating it from the other binding transitions. Moreover, we must require that this second binding transition be irreversible, i.e. $\bar{k}_2 = 0$ and $\bar{k}_4 = 0$. Strictly speaking, however, it is sufficient to require only that $\bar{k}_2 \ll k_3[T]$ and $\bar{k}_4 \ll k_4[T]$ to reduce the velocity to a simple Michaelis-Menten dependence on $[T]$. In this fashion, if the tight-binding is “largely” irreversible, the sigmoidicity can be effectively hidden at low $[T]$ either below the measurement range or within experimental error. This condition corresponds to the requirement that the probability of un-tight binding is very small, and thus this process is negligible. Assuming that $\bar{k}_2 = 0$ and $\bar{k}_4 = 0$, the other three coefficients become

$$a_2 = k_1 k_2 k_3 k_4 k_5$$

$$b_1 = k_1 k_2 k_3 (k_4 + \bar{k}_1) + k_3 k_4 (k_2 + \bar{k}_1)(k_5 + \bar{k}_5)$$

$$b_2 = k_1 k_3 (k_2 + k_3 + k_4 k_5)$$

The above assumption of a linear kinetic scheme requires that the ATPs bind in a time-ordered fashion. However, we can consider a more complicated kinetic scheme in which this requirement is relaxed. For example,

$\begin{align*}
EE \leftrightarrow_{k_1[T]} & \quad TE \leftrightarrow_{k_2} T^*E \\
\downarrow k_1[T] & \quad \uparrow k_2 \quad \uparrow k_4 \quad \uparrow k_4[T] \\
ET \leftrightarrow_{k_1[T]} & \quad TT \leftrightarrow_{k_4} T^*T \\
\uparrow k_2 & \quad \uparrow k_4 & \quad \downarrow k_2 \\
ET^* \leftrightarrow_{k_1[T]} & \quad TT^* \leftrightarrow_{k_2} T^*T^* \rightarrow EE
\end{align*}$

where we have assumed for simplicity that all ATP docking events have a rate of $k_1[T]$ and all tight binding transitions have a rate $k_2$. All reverse rates (not shown) are again denoted with an
overbar. By allowing a second ATP to dock before the first ATP is tightly bound (the TT state), we have relaxed the specific time ordering of binding inherent in the linear scheme. We allow for different forward rates for these additional docking reactions and tight binding reactions, i.e. $k_3[T]$ and $k_4$, and for their reverse rates, $\bar{k}_3$ and $\bar{k}_4$, to facilitate analysis of their effect on the final velocity. Finally, we assume for simplicity that the process ends with an irreversible reset at the end of the cycle.

With the same techniques as above\textsuperscript{7,12}, it can be shown that the velocity for this system is

$$v = d \left( a_2[T]^3 + a_3[T]^3 \right) \left( b_0 + b_1[T] + b_2[T]^3 + b_3[T]^3 \right)$$

which displays the expected sigmoidal dependence on $[T]$. It is instructive to examine the explicit dependence of a few of these coefficients on the rate constants. In particular,

$$b_0 = \bar{k}_1\bar{k}_2 \left( k_3 k_4 (k_4 + \bar{k}_3) + (k_5 + 2\bar{k}_2)(k_4\bar{k}_1 + \bar{k}_4\bar{k}_3 + \bar{k}_1\bar{k}_4) \right)$$

$$b_3 = k_3 k_4^2 \left( k_4 (2k_4 + k_3) + (2\bar{k}_2 + k_3)(2k_4 + \bar{k}_4) \right)$$

$$a_3 = 2k_3 k_4^2 k_2 k_5$$

If we set only two rate constants to zero, namely $\bar{k}_2 = k_3 = 0$, these three coefficients vanish, and the velocity of the system becomes $v = a_2[T]/(b_0 + b_2[T])$. Remarkably, these are the only two rate constants necessary to separate each of the ATP binding events by at least one irreversible transition. Since $\bar{k}_4$ corresponds to essentially the same kinetic reaction as $\bar{k}_2$, it is reasonable to also set $\bar{k}_4 = 0$, in which case

$$a_2 = 2k_1^2 k_2^2 k_5 (k_4 + \bar{k}_3)$$

$$b_1 = 3k_2 k_3 k_4 (k_2 + \bar{k}_1)(k_4 + \bar{k}_3)$$

$$b_2 = 2k_1^2 k_2 (k_2 + 2k_3)(k_4 + \bar{k}_3)$$

Notice that the common factor which contains the now irrelevant rates $\bar{k}_3$ and $k_4$ will cancel from the final expression for the velocity.
The only reaction prohibited by setting $k_3 = 0$ is the loose docking of the second ATP after the first ATP has docked but before it has tightly bound (the $TT$ state.) Physically, this double-docked state could be prohibited in one of two fashions. Either all subunits are initially active for loose binding and the docking of a single ATP temporarily inactivates all other binding pockets until that ATP is committed, or all pockets but one are naturally inactive, and the tight binding of ATP to that pocket activates another pocket for loose binding. We favor the latter case since it involves the least amount of allosteric interaction and because tight binding most likely involves larger conformational changes than docking, conformational changes that could lead to activation of an adjacent binding pocket. It is also important to note that despite the simplifying assumption of many identical rates and only two binding sites, the results derived here can be shown to apply for arbitrary rates for all transitions (calculation not shown) and the conclusions can be easily extended to systems with arbitrary numbers of binding pockets.

To summarize, the sigmoidicity exhibited by these kinetic schemes can be framed in terms of the following general statement: the velocity will display a sigmoidal dependence on substrate concentration when multiple binding transitions are reversibly connected. Conversely, a motor in which each docking transition is separated by an irreversible commitment transition will not display sigmoidicity, regardless of the number of binding sites and the requirement that multiple substrate molecules bind before product formation. This is true even if the rates for binding one substrate molecule differ from the next—that is, that binding at one site facilitates or hinders binding at the subsequent site.

**Integer Step Size Simulations.** An alternative interpretation of the 2.5-bp periodicity observed in the PWD in Figure 3b is that the motor takes integer-sized but non-uniform steps, 2 and 3 bp for example, the average of which would produce a 2.5-bp peak in the PWD. These integer steps can either occur deterministically in time, i.e. a 2-bp step is always followed by a 3-bp step and vice versa, or randomly in time with equal probability.

To rule out these possibilities, we performed the same PWD analysis as described in the methods section with stepping data simulated via these two different models and for a model with uniform 2.5-bp steps. Position versus time data were simulated using a kinetic model where the first dwell time in a set of four dwell times is drawn from a distribution formed from the
convolution of four irreversible transitions with equal rates of 10 s\(^{-1}\) while the additional three dwells are drawn from an exponential distribution with rate 20 s\(^{-1}\), kinetics similar to that observed in Figure 3c and Supplementary Figures 5 and 6. Gaussian distributed noise of either 0.05 bp/Hz\(^{1/2}\) or 0.03 bp/Hz\(^{1/2}\) was added to the simulated data, values similar to that of the traces in Figure 3a, \(\sim 0.04\) bp/Hz\(^{1/2}\). The reported PWD were averaged from 100 simulated packaging traces.

Supplementary Figure 7 shows the results of these simulations. The uniform 2.5-bp step size model reproduces many of the features of the observed PWD, including the larger peaks at 10 and 20 bp and the similar size of the 5, 7.5, 12.5, 15, and 17.5-bp peaks. In contrast, the deterministic model, where steps of 2 and 3 bp alternate in time, predicts that the peaks at 5 and 15 bp will be larger than the non-integer peaks (since 2+3 and 3+2 equal 5), a feature that is not observed in Figure 3b. Moreover, the random model surprisingly fails to produce any of the higher order 2.5-bp peaks beyond the 5-bp peak. Lowering the simulated noise in this model further demonstrates that these higher order peaks never exist (Supplementary Figure 7b), even when the noise is low enough that the individual 1-bp divisions between steps can be seen clearly.

While Supplementary Figure 7 demonstrates that both of the integer step size models produce clear signatures in the PWD which are not observed in Figure 3b, our data cannot rule out the occasional step of different size nor can it rule out variability in the step size on the \(\sim 0.1\)-bp scale. In fact, the decrease in peak height observed for peaks higher than 2.5-bp in Figure 3b may be due to step size variability on this scale, resulting in a modest dephasing of the steps that form these higher order peaks.

**Distortion of DNA.** The step size value of 2.5 bp was obtained by converting the measured contour length change into base pairs assuming the average 3.4 Å/bp rise of B-form DNA\(^{14}\). It is important to note that this conversion is correct irrespective of any distortions in the DNA structure inside the phage capsid or near the motor-DNA contacts, and relies solely on the DNA tether between the trapped beads being predominantly B-form\(^{15}\). Supplementary Figure 8 provides an illustration of this point. Imagine that the motor binds the DNA, distorting a constant number of base pairs in front of the motor contact \(N_d\) and leaving the remaining \(N_b\)
base pairs in undistorted B-form. If the motor steps in increments of $N_{step}$ base pairs then the required conformation change will be $\Delta x' = \beta' N_{step}$, where $\beta'$ is the average rise of the distorted DNA, an unknown quantity. However, once the motor finishes the step, returning to its original mechanochemical state, then by translational symmetry, it must again distort $N_d$ base pairs, resulting in a loss of $N_{step}$ base pairs from the B-form region of the DNA. The net change for the mechanochemical cycle, the value that determines the observed step size, is $N_{step}$ base pairs of B-form DNA and zero base pairs of the distorted DNA. Thus, the presence of distorted DNA does not affect our ability to infer the number of base pairs of DNA packaged from the observed extension changes.

This argument does not require a discrete number of distorted base pairs of DNA nor a discrete transition from distorted DNA to undistorted DNA. It only requires that the distortion propagates over a distance much smaller than the total length of the DNA tether. Such an assumption is valid for two reasons. The first is that our DNA tethers are not torsionally constrained, so any twist applied by the motor—a non-local deformation—will be dissipated from the molecule much faster than our measurement time, ~ 10 ms. The second reason is that structural evidence suggests distortions to DNA are highly local. For example, the crystal structure of a B-Z DNA transition shows that B-form is reestablished only a few base pairs after the highly distorted Z region$^{16}$.
Supplementary Figures and Legends

Supplementary Figure 1: Packaging of DNA by φ29 in Semi-Passive Mode. (a) Calculated contour length of the DNA tether with (b) the associated force and (c) the trap separation as a function of time for a representative trace collected at 100 µM [ATP].
Supplementary Figure 2: General Information for all Reported Data. (a) The number of packaging complexes as a function of [ATP] and under different packaging conditions, low external load (solid circles), high external load (open circles), and low external load but in an 80% D$_2$O buffer (solid triangles). (b) The total length of packaged DNA under each condition. (c) The total number of semi-passive mode segments under each condition. (d) The number of 10-bp increments used for kernel density estimation for the distributions in Figure 2a and Supplementary Figure 5. (e) The fractional error estimated from the internal distance standard between semi-passive mode segments. Error bars represent the standard error of the mean. A negative percent error indicates that the passive mode calculation of the contour length from the observed force and extension underestimates the actual contour length change. The systematic error plotted for 5 µM [ATP] was estimated from the average value observed for all other low external force values, except for the outlier at 25 µM, since the low processivity of the motor at this low [ATP] prevented direct estimation.
Supplementary Figure 3: Pairwise Distributions for All Data and for the Top 50% of All Data. Plotted in black is the average pairwise distribution (PWD) for (a) all data collected at 100 µM [ATP] and low external force, representative of what is observed for all [ATP] at low force, and (b) all data collected at 250 µM [ATP] and high external force. Note that the periodicity observed in the average PWD for the top 50% (red), reported in the main text, is also present in the average PWD for all data (black).
Supplementary Figure 4: The 10-bp Packaging Bursts are Composed of Smaller Steps that are Correlated in Time. (a) Representative packaging traces at low external load show 10-bp packaging bursts are interrupted by brief micro-dwells, highlighted by arrows. Plotted are traces collected at 25 μM (red) and 10 μM (black) at 1.25 kHz in gray and filtered and decimated to 50 Hz in color. Similar events can be observed for all [ATP]. (b) Two dimensional probability distribution for the step sizes and dwell times found using the t-test filter for low external load at 100 μM [ATP], smoothed with a 2x2 boxcar kernel. Similar distributions are observed for all [ATP]. (c) Image plot of the same two dimensional distribution shown in (b) with the boundaries used to define three distinct events, labeled A, B, and C. (d) Conditional step size distributions for steps that precede events in regions A, B, and C, weighted by the probability of each type of event and plotted in blue, green, and red, respectively. The non-conditional step size distribution is shown in black.
Supplementary Figure 5: Packaging in D$_2$O Does Not Affect the 10-bp Burst Size. (a) The average PWD of the top 50% of the packaging data collected at low external force with 250 μM [ATP] in a 80% D$_2$O packaging buffer. The spatial periodicity indicates a burst size of 9.7 ± 0.2 bp (s.d.). (b) Dwell time distribution observed for the 10-bp bursts under these conditions.
Supplementary Figure 6: Additional Analysis of 2.5-bp Steps. (a) Step size histogram for the high force stepping data derived from the t-test analysis (N=5,754.) This distribution has a peak at 2.48 ± 0.03 bp (s.e.m.) and is well described by the sum of four Gaussians equally spaced by a step size \( d \), with equal widths \( w \), and with geometrically decreasing amplitude \( p \). The step size derived from this fit is consistent with a step size inferred from Figure 3b and the amplitude is consistent with the kinetics observed in Figure 3c and the dead-time of the measurement, 20 ms. Error bars are the standard deviation. (b) The dwell time distribution measured for the 2.5-bp steps at high external force, 250 µM [ATP], and an 80% D\(_2\)O packaging buffer is plotted in blue circles with the bi-exponential fit to this distribution in black. Plotted in red is the dwell time distribution observed for the 10-bp bursts under the same packaging conditions but low external force, see Supplementary Figure 5b. The dashed line is this distribution scaled by 0.27, a value consistent with one out of four steps having dwell time distributions drawn from the peaked distribution seen in Supplementary Figure 5b.
Supplementary Figure 7: Pairwise Distance Distributions for Different Step Size Models.

(a) Pairwise distance distributions (PWD) for a model with a uniform 2.5-bp step size (blue), a model in which integer steps of 2 and 3 bp alternate (red), and a model in which 2- or 3-bp steps are taken randomly but with equal probability (green), all calculated from simulated data with Gaussian noise of 0.05 bp/Hz$^{1/2}$. Plotted in (b) are the PWD for the random model with 0.05 bp/Hz$^{1/2}$ noise in green or 0.03 bp/Hz$^{1/2}$ noise in cyan.
Supplementary Figure 8: DNA Distortion By the Motor Does Not Affect the Step Size Measurement. (a) A molecular motor (green) engages the DNA at a specific phosphate (yellow) distorting the DNA out of B-form within the motor and potentially for $N_B$ base pairs in front of the motor. (b) When the motor steps forward, its motion $\Delta x'$ is equal to the number of base pairs per step $N_{\text{step}}$ times the rise of the distorted DNA, an unknown quantity. (c) Upon reengagement of the next phosphate, the same distortion is introduced locally into the DNA, distorting B-form base pairs to replace the $N_{\text{step}}$ distorted base pairs that were translocated. Thus, it is the B-form region of the DNA that is shortened each mechanochemical cycle. The steps observed in (b) and (c) most likely occur simultaneously on the time scale of measurement.
**Supplementary References**


