Supplementary Material for

Kinesin steps do not alternate in size

Adrian N. Fehr, Charles L. Asbury and Steven M. Block
Corresponding author: Steven M. Block, Tel.: 650-724-4046; Fax: 650-723-6132; E-mail: sblock@stanford.edu.

Supplementary Methods

Single-molecule optical trapping assays with recombinant and native kinesin were conducted at saturating ATP levels (2 mM) using protocols described previously (Asbury et al., 2003, *Kinesin moves by an asymmetric hand-over-hand mechanism*. ***Science*** **302**: 2130-2134). Truncated, recombinant conventional kinesin (Kinesin-1) from *Drosophila melanogaster*, DmK401, was expressed and clarified by centrifugation as previously described (*Ibid.*). Clarified lysate was then mixed 1:4 with binding buffer (50 mM $NaPO_4$, 60 mM imidazole, 250 mM NaCl, 1 mM $MgCl_2$, 2 mM DTT, 10 μM ATP, pH 8.0) and incubated on His-binding columns (HisTrap FF Crude; GE Healthcare, Piscataway, NJ) at 4°C for 8 hours. Using an FPLC, columns were washed (wash buffer is identical to binding buffer, but adjusted to pH 6.0) and kinesin protein was eluted with an imidazole gradient (elution buffer is identical to binding buffer contains 0.5 M imidazole and is adjusted to pH 7.0). Kinesin fractions were monodisperse, as judged by SDS-PAGE gels, and stored in 50% glycerol at -20°C until use. Native, conventional squid kinesin was also purified from optic lobes of *Loligo pealii* and used as described (*Ibid.*).

During force-clamped kinesin stepping experiments, position data were acquired at 20 kHz, decimated to 2 kHz and then filtered at the appropriate Nyquist frequency (1 kHz) with a 6-pole Bessel filter. Under our assay conditions (temperature 21.2° C), the mean kinesin speeds at the retarding forces used (~150 nm/s) imply an average dwell time of ~50 ms. Hence, the vast majority of kinesin dwell intervals could be clearly resolved beyond the thermal noise (see Fig. S1). These events were identified and scored using an automated step-finding algorithm written in Igor Pro 5.0 (WaveMetrics, Lake Oswego, OR). All stepping transitions identified in this fashion were subsequently confirmed by eye, and adjusted if necessary. Note that for the determination of stepping *distances* (in contrast to stepping *times*), the measurement of records is fairly robust against any small errors made in the precise determination of the locations of stepping transitions, because the positional average for each dwell interval is dominated by the

majority of data found in the plateau region for each step, and not by the transient changes that occur at the step boundaries. To be included for analysis, we required a minimum duration of 4 ms for a dwell interval. For the purpose of this study, we assumed that any brief, 16-nm displacements simply corresponded to back-to-back 8-nm steps that were unresolved, and therefore were disregarded. Such events were rare (<5% of all steps). Details of our step-finding algorithm are found in the Appendix to the Supplemental Material. For a recent discussion of the performance of step-finding algorithms, please refer to (Carter, B.C., Vershinin, M., and S.P. Gross, *A comparison of step-detection methods: how well can you do?* ***Biophys. J.***, 2007, ePub ahead of publication doi:10.1529/biophysj.107.110601). The code to our algorithm, and to other routines produced by our laboratory for this work, is available upon request.

Fits to models consisting of either one or two Gaussian distributions were compared statistically by means of the *F*-test, which is based on the ratio of reduced chi-squared values (i.e., chi-squared divided by the number of degrees of freedom) for the two candidate distributions (see, for example, Bevington, P. R., and D. K. Robinson. 1992. *Data Reduction and Error Analysis for the Physical Sciences*. 2nd Edn. McGraw-Hill, New York.). Implicit in the use of the *F*-test is the assumption that experimental deviations from the parent (model) function follow the usual chi-square distribution, which in turn requires that errors at data points be normally distributed. Counting errors computed for individual histograms bins were taken to be statistical errors (i.e., equal to $\sqrt{N}$, where N is the number of counts). The distributions of such counting errors are very well approximated by the normal distribution whenever $N \geq 10$. Bins with $N < 10$ counts were excluded from fits (Legend to Fig. 2).

Supplementary Data 1: Phase assignment errors. For a kinesin molecule that steps with alternating stochastic dwell times that differ by a factor of ~5, on average, there is a finite probability that in records containing small numbers of steps, the "slow" phase will appear to be faster than the "fast" phase, due to limited sampling. To test the effect of this sampling error on our measurements, we created simulated stepping records *in silico* with exponentially-distributed numbers of steps (mean = 5 steps) and sequential dwell times drawn from exponential distributions with different time constants, similar to those values previously found in actual limping kinesin records (Asbury et al., 2003, ***Science*** **302**: 2130-2134). We found that the stochastic nature of stepping caused relatively few mistakes in assignment: 350 errors per

50,000 records, or 0.7%. However, this represents a best-case scenario, because it remains possible that a small percentage of steps are not resolved in a given record, which would cause the phase to be lost, and thereby increase the error rate. Missing a step effectively lowers the limp factor for that record however. Therefore, we excluded runs with $L < 5$ and redid the simulations, and found that the error rate was thereby reduced to ~2 per 50,000 records, or 0.004%. Setting the threshold in this manner therefore reduces the possibility of missed phase assignments to negligible levels.

Supplementary Data 2: Stepping controls. To determine if these experiments supplied sufficient resolving power to distinguish alternating steps separated by ~2 nm, we performed a hardware-based simulation of kinesin stepping. Beads carrying limiting numbers of kinesin molecules (diluted to a limit where only a single molecule would bind) were immobilized on microtubules using the non-hydrolyzable ATP analog, AMP-PNP. Records of bead position were then collected using our standard force-clamp software as the piezoelectric stage position was moved in alternating 6.9- and 9.1-nm increments under computer control (in effect, pulling the kinesin/bead across the detection region as a kinesin molecule normally would), with sequential dwell times drawn from exponential distributions with two different time constants (Fig. S2 *A*). Position data were analyzed in a manner identical to normal kinesin stepping data. After separating steps by phase, we found that the slow and fast components corresponded to steps measuring 7.1 ± 0.1 nm ($N = 419$) and 9.3 ± 0.1 nm ($N = 420$), respectively (Fig. S2 *B*), consistent with the input parameters. We also pooled all step sizes and fitted these data with both a single Gaussian (with fixed mean of 8.2 nm) and a double-Gaussian (with fixed means of 6.9 and 9.1 nm). We found that the $\chi^2_\nu = 9.8$; $N = 7$; $P \sim 0.001$ for the Gaussian and $\chi^2_\nu = 1.9$; $N = 7$; $P \sim 0.1$ for two Gaussians ($F = 5.2$; $P = 0.05$), again consistent with steps being drawn from a parent distribution with two step sizes, not one.

Figure S1: Representative stepping records for three DmK401 molecules. Limping motion is apparent as alternating slow and fast dwell intervals, scored by an automated algorithm and colored blue and red, respectively. Top record: $L = 7.4$ with step sizes of 9.1 ± 1.5 nm (slow phase) and 9.0 ± 2.0 nm (fast phase); mean ± s.d. Middle: $L = 17.9$ with step sizes of 7.9 ± 1.3 nm (slow phase) and 8.8 ± 1.5 nm (fast phase). Bottom: $L = 22.0$ with step sizes of 8.5 ± 2.2 nm (slow phase) and 8.8 ± 1.5 nm (fast phase).

Figure S2: 7- and 9-nm stepping data. Mechanical simulation of an alternating 7- and 9-nm stepper. **(a)** Position records of a bead tethered to a microtubule by a single immobilized kinesin molecule as the computer-controlled stage executed 7- and 9-nm steps with times drawn from exponential distributions, thereby simulating kinesin motion with alternating step sizes and experimental noise levels comparable to an experiment with functional kinesin. $F \sim -4$ pN. **(b)** Experimental distribution of step sizes separated by phase. Slow step size = 7.1 ± 0.1 nm and fast step size = 9.3 ± 0.1 nm. Arithmetic mean ± s.e.m. **(c)** Distribution of all step sizes showing a single-Gaussian fit (blue) and double-Gaussian fit (red). Only bins with ≥ 10 counts were included in fits.

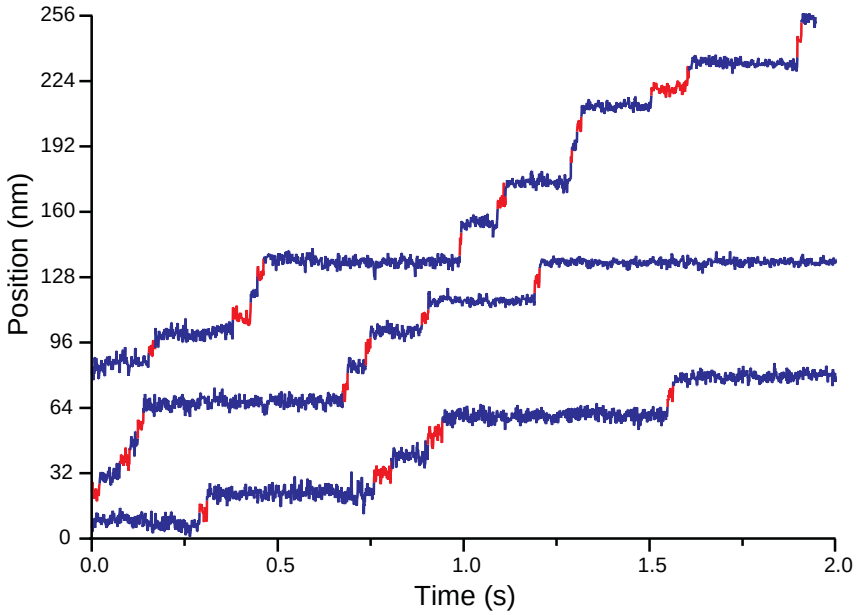

Figure S1
Fehr, Asbury and Block, 2007

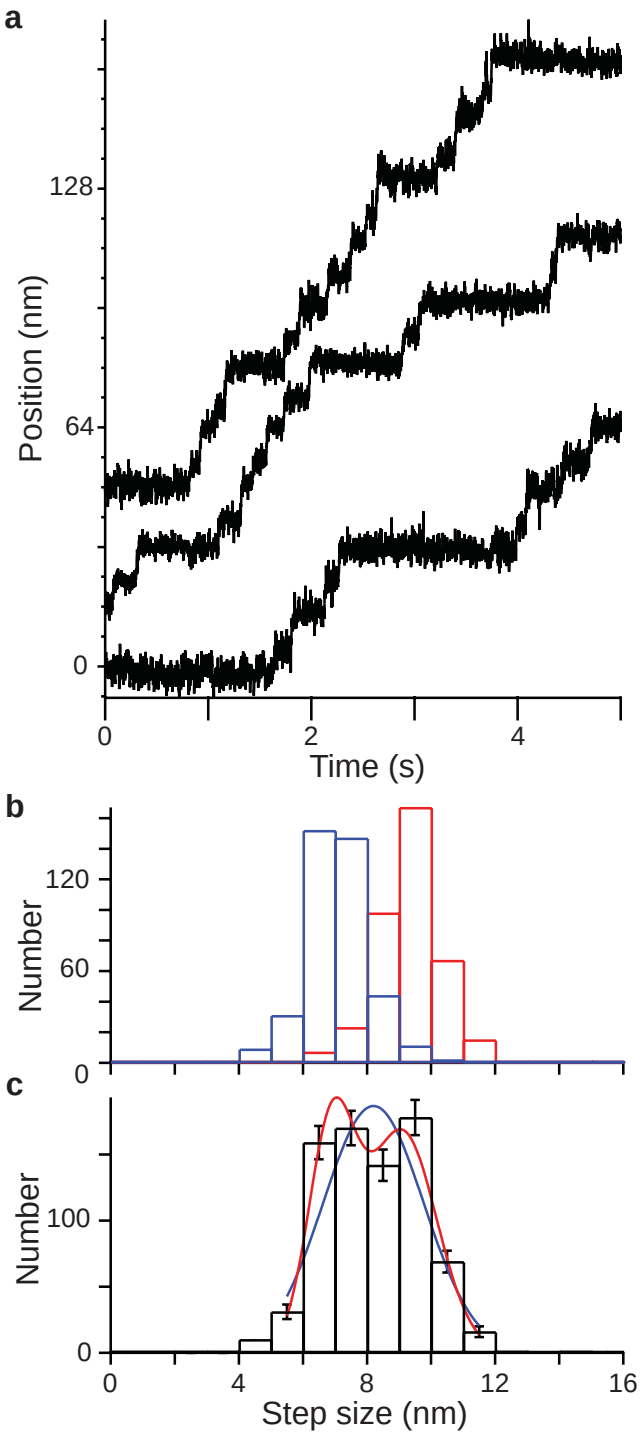

Figure S2
Fehr, Asbury and Block, 2007

APPENDIX

An automated step-finding algorithm implemented in IgorPro 5.0.

Input data: a computer record of position vs. time (typically acquired at 20 kHz, decimated to 2 kHz, and filtered at 1 kHz).

Output data: a list of all successive step sizes.

Key assumption: a nominal (approximate) step size, taken to be 8.2 nm for kinesin (the exact value is not critical).

Algorithm:

1. Smooth the input record of position vs. time using a lowpass boxcar filter (at 0.1-0.5 of the record data rate).
2. Use IgorPro's *Histogram* operation to determine the distribution of positions, using a fine bin size (≤ 1 nm).
3. Use IgorPro's *Derivative* function twice to compute the 2nd derivative (negative peaks reflect dwell positions).
4. Use IgorPro's *FindPeak* operation, with a selected threshold, to create a list of all negative peak locations (i.e., dwell positions).
5. Subtract adjacent dwell positions to generate a list of successive separation distances between events. (Note: if some peaks are missed in step (4), this may return some values that are near integral multiples of the nominal step size.)
6. Using information from the lists created in steps (4) and (5), go back and use IgorPro's *FindLevels* operation, which identifies levels crossing beyond a settable threshold, to identify the final list of times of stepping transitions in the original input record, as follows.
    a. Go systematically through the list generated in step (5), one transition distance (step) at a time.
        i. If this distance is roughly equal to the nominal step size, then set a threshold at the midpoint between the two levels generated in step (4) and apply *FindLevels* with this threshold to identify the transition time.
        ii. If this distance is roughly equal to twice the nominal step size, a (brief) step has likely been missed. In this case, establish two thresholds at ¼ and ¾ of the distance between the levels generated in step (4). Apply *FindLevels* twice, using these thresholds, to identify two nearby transition times.
        iii. If the distance exceeds twice the nominal step size, then flag the event and alert the user not to use this record.
        iv. Verify the transition. To avoid false level crossings (e.g., due to occasional, brief noise spikes in the data), take an average of the input position data on either side of each identified transition to confirm a sustained level change (typically, for 10 ms on either side). If these values differ significantly, the step is verified. If not, then use *FindLevels* with the same threshold to continue searching the input data for the actual transition.

7. Final sanity check: estimate the number of stepping events expected, by dividing the total distance traveled divided by the fundamental step size. Compare this number with the number of steps identified by the algorithm. Alert the user if these numbers differ.
8. Produce final output: find all step sizes within the input record based on the locations of transition times.
    a. Go through the input data record using the list of times generated in step (6), and compute the arithmetic mean for all position values contained within each segment between adjacent, identified transitions.
    b. Generate the list of successive differences from the values identified in step (8a). These are the step sizes.