PLANT: A Method for Detecting Changes of Slope in Noisy Trajectories

Alberto Sosa-Costa[1], Izabela K. Piechocka[1], Lucia Gardini[2,3], Francesco S. Pavone[2,3,4], Marco Capitanio[2,4], Maria F. Garcia-Parajo[1,5], Carlo Manzo[1,6,*]

[1]ICFO-Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, Castelldefels, Barcelona, Spain; [2]LENS - European Laboratory for Non-linear Spectroscopy, Sesto Fiorentino, Italy; [3]National Institute of Optics–National Research Council, Florence, Italy; [4]Department of Physics and Astronomy, University of Florence, Sesto Fiorentino, Italy; [5]ICREA, Barcelona, Spain; and [6]Universitat de Vic - Universitat Central de Catalunya, Vic, Spain.

ABSTRACT

Time traces obtained from a variety of biophysical experiments contain valuable information on underlying processes occuring at the molecular level. Accurate quantification of these data can help to understand the details of the complex dynamics of biological systems. Here, we describe PLANT (Piecewise Linear Approximation of Noisy Trajectories), a segmentation algorithm that allows the reconstruction of time trace data with constant noise as consecutive straight lines, from which changes of slopes and their respective durations can be extracted. We present a general description of the algorithm and perform extensive simulations to characterize its strengths and limitations, providing a rationale for the performance of the algorithm in the different conditions tested. We further apply the algorithm to experimental data obtained from tracking the centroid position of lymphocytes migrating under the effect of a laminar flow and from single myosin molecules interacting with actin in a dual-trap force-clamp configuration.

INTRODUCTION

The output of several biophysics experiments, such as DNA elongation via magnetic and optical tweezers (1–4), myosin motor motility (5), and cell tracking under shear flow (6, 7) consists of time trace curves. Many of these curves show alternate regions of pauses and linear increments usually associated with different events taking place at the molecular level (6, 7). The identification of such regions and the associated physical parameters such as segment duration and slope are essential to better understand the underlying dynamics of the ongoing processes. However, the noise produced by thermal fluctuations affects these data and limits their accurate quantification. Thus, the automatic detection of the different features of the time trace curves remains a challenging computational problem.

Several approaches have been developed for the automatic segmentation of time traces (8–11), mainly for the dimensionality reduction in data mining (12). One of the most common approaches is the approximation of the original data by different straight lines, usually referred to as Piecewise Linear Approximation (PLA). Given a time series, this approach consists in constructing a piecewise linear function that produces the best representation, either using a limited number of segments, or not exceeding a user-specified threshold for the error (8). Generally speaking, the segmentation algorithms based on the PLA can be grouped into three different categories: sliding window, top-down, and bottom-up algorithms (8). In the sliding window algorithms, a segment is grown until the error after adding a new point to the segment exceeds some specified threshold. Although this type of approach is ideal to analyze the data while being generated (i.e. online segmentation), it tends to overestimate the number of segments (13). On the other hand, the top-down type of algorithms work by recursively splitting the entire trace until certain error threshold is reached. Last, bottom-up algorithms start by arbitrarily dividing the trace in multiple segments and recursively merging the two more similar adjacent segments. The process is stopped when some criterion fixed by the user is met. Although, in general, top-down and bottom-up algorithms have shown to perform better than the sliding window algorithm (8), there are several variations of these three methods, each of them usually adapted to perform better on specific type of data (14–17).

A crucial element of the PLA consists in determining the breakpoints between segments of different slopes. This is related to a more general problem in statistics, concerning with the determination of the actual states of a system whose

*Correspondence: carlo.manzo@uvic.cat

observables are corrupted by noise (18). This problem can be approached by the so-called change-point analysis (19), first introduced by E. S. Page (20, 21), which assesses whether a statistically significant change in a given observable occurs at some point within the time trace (22, 23). Algorithms based on change-point analysis have been developed to segment time traces originated by a variety of experiments, including biophysical and single-particle experiments (22, 24–29).

Here, we describe a bottom-up type of PLA algorithm that uses change-point analysis to robustly reconstruct time traces with constant noise. Unlike conventional bottom-up algorithms, in a first step we find the points where there is a potential variation of the slope and then we recursively merge the adjacent segments. We follow a likelihood-based approach and therefore we do not impose any condition on the number of segments or on the residual error. While most of the change-point algorithms developed for the segmentation of biophysical data were meant to detect abrupt changes of the mean value or the variance in time traces, our algorithm deals with the detection of change points in a multiple linear regression model (30). A similar approach was recently described and used to detect change in DNA looping dynamics during replication (28) or changes in diffusion and/or velocity in single particle trajectories (29). These works followed the binary segmentation scheme first described in (24) and used a maximum likelihood ratio test, whereas our method relies on a different segmentation approach to find putative change points and later applies a Fisher test to eliminate spurious identifications. This scheme allows to evaluate all the points within a trace as potential change points, and to further refine them according to a hierarchical statistical criterion. We tested our algorithm on a variety of conditions, including different noises, slopes, and segment lengths and provide reasonable explanations about the observed results. This analysis shows performance equal or higher than other published methods. Although strictly speaking our method is less generally applicable than others due do the constraint of a nearly constant noise along the trace, this condition is met - at least locally - in several types of single molecule experiments, as those discussed in the last sections of this paper. Finally, we validated our algorithm on time traces experimentally obtained from i) tracking the centroid of lymphocytes moving on endothelial cells (EC) under the presence of a constant shear flow and ii) ultrafast force-clamp spectroscopy experiments aimed at quantifying force-dependent kinetics of myosin from skeletal muscle.

MATERIALS AND METHODS

Algorithm description

The flow diagram of the algorithm is depicted in Fig. 1 A. We consider an input data consisting in a trajectory in which an observable y is sampled at N discrete time points t_i. The working hypothesis is that the trajectory $(t_i, y(t_i))$ is composed by a sequence of K adjacent linear segments, bounded by $K+1$ change points - including the trajectory boundaries - occurring for values of the index $i = cp_k$ and corrupted with Gaussian noise having constant variance σ^2 (Fig. 1 B). Given a minimum segment length N_{MIN}, an estimation of the noise σ^2 and a confidence level $1-\alpha$, the objective of the algorithm is to efficiently and precisely determine the change points cp_k (Fig. 1 C), and thus reconstruct the original trace to finally be able to measure the segments duration and slope (Fig. 1 D). The algorithm is composed by two main sequential blocks, the first of which (*WINLINFIT*) recursively selects a set of candidate change points associated to a potential variation in the slope by means of least square fitting and likelihood evaluation, whereas the second (*MERGING*) refines the number of change points by hierarchically executing an hypothesis test on adjacent segments with a confidence level $1-\alpha$. Analysis of time traces of various length shows that the time complexity of the algorithm scales as $O(n^2)$, and it typically takes $\sim$ 0.5s to analyze a 100-points long trace on a MacBook Pro with 2.2 GHz Intel Core i7 processor and 16Gb RAM. For long data streams, segmentation of the trace in overlapping segments and sequential application of the algorithm can be used to reduce computational time. Alternatively, parallelization schemes such as those described in (31, 32) can also be used to improve scalability in runtime and computational memory.

WINLINFIT

The *WINLINFIT* module operates on the input trajectory to provide an initial estimation for the change points cp_k. Under the hypothesis that the trajectory is a concatenation of linear segments in the presence of Gaussian noise, performing a least square fitting over each of these segments with a linear model $f(t_i) = a_k + b_k t_i$ allows to calculate the estimators $\hat{a}_k$ and $\hat{b}_k$ (33). Therefore, by applying the principle of maximum likelihood, we obtain that

$$s_k^2 = \frac{\sum_{i=cp_k}^{cp_{k+1}} (a_k t_i + b_k - y_i)^2}{cp_{k+1} - cp_k + 1}. \quad (1)$$

The quantity s_k^2 can also be interpreted as the variance of a sample of length $N_k = cp_{k+1} - cp_k + 1$ and its expectation value provides the (biased) estimator of the population variance σ^2. It has been demonstrated that s_k^2 has a Pearson type III probability distribution (34):

$$p\left(s_k^2, N_k|\sigma^2\right) = \frac{\left(\frac{N_k}{2\sigma^2}\right)^{\frac{N_k-1}{2}} \left(s_k^2\right)^{\frac{N_k-3}{2}} e^{-\frac{N_k s_k^2}{2\sigma^2}}}{\Gamma\left(\frac{N_k-1}{2}\right)}, \quad (2)$$

where $\Gamma(\cdot)$ represents the Gamma function.

Although for the trajectories under examination neither the number $K+1$ nor the location of the cp_k are known, in many cases it is possible to obtain an estimation of σ^2, either from control experiments or from the trajectory itself. An example of such kind of procedure is provided in Supporting Material. We can split the trajectory in all the possible combinations of segments having length $N_k \geq N_{MIN}$, with $N_{MIN} \geq 3$ imposed by the condition of having a number of points larger than the degrees of freedom for a linear fit. The total number of segments is calculated as

$$\begin{aligned}\sum_{N_k=N_{MIN}}^{N} (N_k - N_{MIN} + 1) &= \cdots \\ &= \tfrac{1}{2}\left(N_{MIN} - N - 2\right)\left(N_{MIN} - N - 1\right). \end{aligned} \quad (3)$$

We thus perform a linear fit on each segment, building a matrix of variances. For the sake of simplicity, we will now label the segments with the couple of indexes corresponding to their bounding cp's, such that $N_k = N_{ij} = j - i + 1$. With this notation we obtain

$$s_{ij}^2 = \frac{\sum_{l=i}^{j} \left(\hat{a}_{ij} t_l + \hat{b}_{ij} - y_l\right)^2}{j - i + 1} \quad (4)$$

and evaluate the corresponding likelihood $p_{ij} = p\left(s_{ij}^2, N_{ij} | \sigma^2\right)$ by means of Eq. 2. The segments fully contained in a linear fragment of the trajectory will provide higher value of the probability, whereas those spanning over two or more linear regions will display larger deviations from the expected variance and thus lower p_{ij}'s. We would like to point out that here we neglect the statistical multiplicity associated with change point selection, which would require a more complicated treatment. Although this simplification limits somewhat the change point localization precision, it largely simplifies the calculations. We then apply a recursive routine, by means of which the segment with the maximum p_{ij} value is selected and the coordinates of its endpoints (t_i and t_j) are saved as potential change points. All the elements of the matrix p_{ij} corresponding to segments having one end point falling within the interval (t_i, t_j) are removed and excluded from the subsequent analysis. The process is repeated until the whole trajectory is segmented in a series of contiguous fragments with change points cp.

MERGING

Once the potential cp's are obtained from the *WINLINFIT* step (empty and full circles in Fig. 1 C), the *MERGING* block statistically refines their occurrence by hierarchically applying a Fisher test with confidence level $1-\alpha$. The *MERGING* block determines whether some of the change points correspond to a false positive detection (empty circles in Fig. 1 C) whose removal would allow the adjacent segments to merge. For each pair of adjacent segment (i, j) and (j, k), we test the null hypothesis H_0 that the segments belong to the same linear region (i, k) and there is no change point j between them, versus the alternative hypothesis H_1, i.e. that the segments do correspond to different linear behaviors. In the first case (H_0), we perform a linear fit over the whole region (i, k), whereas in the second case (H_1) two different linear fits are executed. The significance level α represents a Type-I error, i.e. the probability of rejecting the null hypothesis when no change point are present. Then, the Fisher statistic is computed as (35):

$$\begin{aligned} F &= \frac{RSS_{H0} - RSS_{H1}}{RSS_{H1}} \left(\frac{N_{ik} - p_{H1}}{p_{H1} - p_{H0}}\right) = \cdots \\ &= \left(\frac{(k-i+1)s_{ik}^2}{(j-i+1)s_{ij}^2 + (k-j+1)s_{jk}^2} - 1\right)\left(\frac{k-i-3}{2}\right), \end{aligned} \quad (5)$$

where RSS_{Hi} and p_{Hi} are the squared sum of residuals and the number of degrees of freedom associated to the two hypothesis $i = 0, 1$, respectively. We then execute the hypothesis test on the pair of segments showing the minimum F value (higher probability to be merged). The value of F is compared with the critical value of the Fisher-Snedecor distribution $\mathcal{F}_\alpha$ (35) with level of confidence $1-\alpha$. In the case the hypothesis H_0 can not be rejected, the segments are merged and the Fisher statistic recalculated for the updated list of cp's. The procedure is repeated as long as the segments pair with the minimum F verifies the null hypothesis condition $\min\{F\} \leq \mathcal{F}_\alpha$.

For multidimensional trajectories, under the assumption that the noise level is the same in all the dimensions, the algorithm can be extended in a straightforward fashion by calculating the total s_k^2 and the squared sums of residuals as a sum over all the dimensions and adjusting the number of degree of freedom accordingly.

RESULTS AND DISCUSSION

General considerations

To first investigate the performance of our algorithm in identifying a change point, we used numerical simulations. For this, we used the scheme depicted in Fig. S1 A in the Supporting Material, made of trajectories composed by two different linear segments with lengths L_1 and L_2, slopes m_1 and m_2 and noise σ. For each set of parameters, 500 traces in total were simulated and analyzed with the PLANT algorithm using $\alpha = 0.05$. The noise standard deviation was automatically estimated by means of the algorithm routine described in Supporting Material. In order to quantify the algorithm performance, we evaluated several figures of merit. An important parameter is the power of detection, i.e. the probability of detecting a change in slope when it is known there is a change within the trajectory. In Fig. 2 A-C, we show that the power of detection depends in a sigmoidal

fashion on the noise σ. Fig. 2 A reports the dependence on the segment length, showing that traces formed by segments of the same length with different slope values but same slope difference $|m_2 - m_1|$ display analogous behavior. In addition, larger difference $|m_2 - m_1|$ (Fig. 2 B) produced a higher power of detection. We also investigated traces composed by segments of different length, finding that the power of detection is actually a function of the segment length through the quantity $N^* = 2N_1N_2/(N_1 + N_2)$ (Fig. 2 B). As discussed in Supporting Material, the dependence on these parameters can be expressed in terms of a scaled variable ξ defined in Eq. S6, for which all the curves collapse into the same one (Fig. 2 C). Next, we calculated the precision in the localization of the cp, defined as the standard deviation of the distribution of the detected cp position minus the "true" cp position. Fig. 2 D shows that this quantity depends on the noise, the slope and the difference between slopes. We also evaluated the error on the detected slopes, as displayed in Fig. 2 E. Δm_i is found to be independent on the segment slopes and to scale linearly with the noise. Last, in Fig. 2 F we report the false positive rate, i.e. the probability of detecting a change point where it actually there is no change point. The data show no clear trend with segment slope and noise, due to the fact that the statistical tests performed by the algorithm explicitly take into account the presence of the noise and its value. The false positive rate shows a dependence on the trace length ($\sim N^{1/2}$, Fig. S2 A), with values lower that the false positive rate specified through the Type-I error level $alpha = 0.05$ ($\lesssim 0.1$ for $L \lesssim 500$)) ensuring high specificity and robustness against spurious detection caused by noise-induced fluctuation. In order to allow the comparison with other works that report the false positive rate per point within a trajectory, we would like to point out that in Fig. 2 F we plot the false positive rate per trajectory. Normalizing by the trajectory length N, we obtain much lower values for the false positive rate. For the cases shown in Fig. 2 F, the average false positive rates per point is of the order of 10^{-4} already for relatively short trajectories (< 200 points); as also reported by others (29), this value further decreases as a function of the trajectory length.

The actual scenario to which the algorithm is meant to be applied is more complex than the simulation scheme depicted in Fig. S1 A. This is because each segment of the trace is embedded within two others with different slopes and durations. Moreover, both a missed change point and a false positive detection will affect the determination of the parameters of the neighboring segments in a complex way, that ultimately depends on the distribution of slopes and durations of the trace. Therefore, the specific performance of the algorithm must be evaluated in every specific case by means of simulations. In order to provide a flavour of the ability of the algorithm to correctly identify a segment in a more realistic scenario, we also used a more complex simulation scheme composed by three neighbouring segments (Fig. S1 B). The results of these simulations are described in Supporting Material (Fig. S2).

Application of the algorithm to the analysis of trajectories from cell tracking experiments

We applied our algorithm to analyze trajectories describing the motion of T-cells over endothelial cells under shear flow (Fig. S3). Typical time traces of the T-cell centroid coordinates in the direction perpendicular (x) and parallel (y) to the flow are shown in Fig. 3 A. Lymphocytes display heterogeneous motions, which at the experimental time resolution appear as composed by a concatenation of segments. Each segment has an approximately constant velocity but we observe large velocity variations between different segments, including reversing the velocity by moving against the flow direction ($v_y < 0$). Therefore, we applied our algorithm to a set of experimentally collected time traces to automatically characterize the different features in the cells motion. The result of the algorithm applied to the y coordinate of the cell centroid position is shown in Fig. 3 A, where the shaded gray regions correspond to the detected changes of slope and the dashed line represents the piecewise reconstructed trace, which has been shifted up for clarity. This analysis was applied to 50 traces corresponding to different cells with duration of 400-600 s, and the distributions of segment lengths (durations) and slopes (velocities) obtained from the analysis are represented as histograms in Fig. 3 B-C. To verify whether the motion along or against the flow had different characteristics, the data corresponding to the duration of the segments (Fig. 3 B) were split based on the direction of the velocity component along the flow direction v_y, which is reported in panel Fig. 3 C. Both distributions show an asymmetric shape, reflecting the fact that the T-cells preferentially move along the flow. In spite of this preference, molecular interactions established by cells allow them to sustain the motion against the shear flow with non-negligible speed. We attempted to obtain a more quantitative insight from these data to link the observed features to a molecular model of cell migration (36, 37). To this aim, we performed stochastic simulations in which we generated synthetic traces with *a priori* known values of velocities and durations to evaluate the impact of the algorithm on the determination of these parameters (see Supporting Material for a full description). By means of this approach, we could find a set of parameters that could nicely reproduce our experimental results (black lines in Fig. 3 B-C). This allowed us to estimate the real parameters of the distributions, before the limitations introduced by instrumental resolution and algorithm limited performance. The results showed that T-cells display distributions compatible with simulations in which the movement along the flow direction occurs with an average speed of 39 ± 6 nm/s, which decreases to 29 ± 5 nm/s when moving against the flow. The typical duration of a segment travelled

at nearly constant speed is of 35 ± 6 s, independently of the direction, producing an average "step" of ~ 1.4 μm in the positive and ~ 1 μm in the negative direction. In addition, the cells spend 2/3 of the time travelling along the flow. These results provide us a first test of the ability of our algorithm to extract valuable information from noisy trajectories obtained from cell tracking experiments. The values of cell velocity and segment duration that can be obtained by means of our algorithm thus constitute relevant microscopic parameters, which can be used to validate molecular models of cell migration (36, 37) and thus better understand the role played by molecular interactions.

Application of the algorithm to the analysis of trajectories from ultrafast force-clamp spectroscopy

We also applied the PLANT algorithm to data from ultrafast force clamp spectroscopy experiments on actin-myosin interaction. Here, an actin filament is stretched between two optically trapped beads and set in contact with a third bead fixed onto the coverslip, which is coated with myosin II at single molecule concentration (Fig. S4). In these experiments, a constant force F is applied to the bead-actin-bead "dumbbell" and alternated back and forth, thus producing a saw-tooth like time trace with regions of positive (negative) slope interrupted by flat regions corresponding to binding events of myosin to actin (Fig. 4 A). The algorithm could efficiently locate the flat segments along the traces. In agreement with previous results (3), the velocity distribution returned by our algorithm clearly showed the existence of two populations (gray histogram in Fig. 4 B, one with mean velocity close to zero and corresponding to the binding events and another peaked around the velocity caused by the application of the force on the dumbbell ($v = F/\gamma$, where γ is the viscous drag coefficient of the dumbbell; unbound state). However, probably due to the presence of correlated noise, the algorithm produced a large number of very short segments with slightly different slopes. To refine the analysis and better assign the events to these two states, we used the same criterion described in the original manuscript (3), i.e. a threshold velocity corresponding to the weighted average of the peak velocities (dashed black line in Fig. 4 B). This procedure groups together adjacent segments with slopes compatible with the occurrence of the same state, producing a clear narrowing of the velocity distribution (open bars of the histogram in Fig. 4 B). As shown in Fig. 4 A, this procedure produces a reconstruction of the flat regions with high fidelity and allows us to detect binding events as short as 100 μs, corresponding to 20 data points. The algorithm retrieved 126 binding events on the analyzed portion of the trace. The cumulative probability distribution of the duration of these events is displayed in Fig. 4 C. In agreement with previous analysis obtained by means of a different method (3), the distribution of bound state duration was found compatible with a 2-exponential function. From the fitting we obtained two different rates $k_1 = 103 \pm 5\ \mathrm{s}^{-1}$ and $k_2 = (4 \pm 3) \times 10^3\ \mathrm{s}^{-1}$. Although the faster rate has a rather large error since it corresponds to durations of the same order of magnitude as the algorithm dead time, the rate values are fully consistent with earlier results and provide higher time resolution as compared with previous methods (3). As such, the algorithm provides an efficient automated tool to extract rapid kinetics from time traces generated by this type of single molecule experiments, including, for example, those measuring the sequence-dependent interaction of single transcription factors along DNA (38).

CONCLUSIONS

We introduced PLANT, a bottom-up type of algorithm for detecting changes of slopes in time traces affected by Gaussian noise. Unlike others algorithms, PLANT does not impose any condition on the number of segments composing the trajectory, and includes a routine for the estimation of unknown noise from the data. By means of extensive numerical simulations, we have shown that it robustly reconstructs noisy trajectories with a power of detection of change points that depends on the length of the segment, the noise and the slope difference. The algorithm is also provided as a flexible and easy to use graphic user interface. Finally, we applied the algorithm on a set of experimental data obtained by tracking the position of T-cells and single myosin molecule interacting with actin under force. These results provide a first test of the ability of our algorithm to extract valuable information from noisy trajectories.

PROGRAM AVAILABILITY

The program is available on the GitHub repository hosting service at http://github.com/cmanzo/PLANT and includes the program files, documentation and demo data.

AUTHOR CONTRIBUTIONS

A.S.-C. and C.M. designed and developed the software. A.S.-C. carried out the simulations and analyzed the data. A.S.-C., I.K.P., and L.G. performed experiments. F.S.P., M.C., M.F.G.-P., and C.M. supervised the research. A.S.-C., M.F.G.-P., and C.M. wrote the article. All authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the European Commission (FP7-ICT-2011-7, grant number

288263), Erasmus Mundus Doctorate Program Europhotonics (grant number 159224-1-2009-1-FR-ERA MUNDUS-EMJD), Spanish Ministry of Economy and Competitiveness ("Severo Ochoa" Programme for Centres of Excellence in Research & Development SEV-2015-0522, and FIS2014-56107-R grants), Generalitat de Catalunya through the CERCA program, Italian Ministry of University and Research (FIRB "Futuro in Ricerca" 2013 grant n. RBFR13V4M2 and Flagship Project NANOMAX), Fundació Privada CELLEX (Barcelona), Ente Cassa di Risparmio di Firenze, Human Frontier Science Program (GA RGP0027/2012) and LaserLab Europe 4 (GA 654148). C.M. acknowledges funding from the Spanish Ministry of Economy and Competitiveness (MINECO) and the European Social Fund (ESF) through the Ramón y Cajal program 2015 (RYC-2015-17896).

SUPPORTING CITATIONS

References (39–52) appear in the Supporting Material.

REFERENCES

1. Haber, C., and D. Wirtz, 2000. Magnetic tweezers for DNA micromanipulation. Rev. Sci. Instrum. 71:4561–4570.
2. Yan, J., D. Skoko, and J. F. Marko, 2004. Near-field-magnetic-tweezer manipulation of single DNA molecules. Phys. Rev. E 70:011905.
3. Capitanio, M., M. Canepari, M. Maffei, D. Beneventi, C. Monico, F. Vanzi, R. Bottinelli, and F. S. Pavone, 2012. Ultrafast force-clamp spectroscopy of single molecules reveals load dependence of myosin working stroke. Nat. Methods 9:1013–1019.
4. Capitanio, M., and F. S. Pavone, 2013. Interrogating biology with force: single molecule high-resolution measurements with optical tweezers. Biophys. J. 105:1293–1303.
5. Yildiz, A., J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman, and P. R. Selvin, 2003. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. Science 300:2061–2065.
6. Décavé, E., D. Rieu, J. Dalous, S. Fache, Y. Bréchet, B. Fourcade, M. Satre, and F. Bruckert, 2003. Shear flow-induced motility of Dictyostelium discoideum cells on solid substrate. J. Cell Sci. 116:4331–4343.
7. Thomas, W., M. Forero, O. Yakovenko, L. Nilsson, P. Vicini, E. Sokurenko, and V. Vogel, 2006. Catch-bond model derived from allostery explains force-activated bacterial adhesion. Biophys. J. 90:753–764.
8. Keogh, E., S. Chu, D. Hart, and M. Pazzani, 2001. An online algorithm for segmenting time series. In Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on. IEEE, 289–296.
9. Hamann, B., and J.-L. Chen, 1994. Data point selection for piecewise linear curve approximation. Comput. Aided Geom. Des. 11:289–301.
10. Popivanov, I., and R. J. Miller, 2002. Similarity search over time-series data using wavelets. In Data Engineering, 2002. Proceedings. 18th International Conference on. IEEE, 212–221.
11. Perng, C.-S., H. Wang, S. R. Zhang, and D. S. Parker, 2000. Landmarks: a new model for similarity-based pattern querying in time series databases. In Data Engineering, 2000. Proceedings. 16th International Conference on. IEEE, 33–42.
12. Keogh, E., K. Chakrabarti, M. Pazzani, and S. Mehrotra, 2001. Dimensionality reduction for fast similarity search in large time series databases. Knowledge and information Systems 3:263–286.
13. Shatkay, H., and S. B. Zdonik, 1996. Approximate queries and representations for large data sequences. In Data Engineering, 1996. Proceedings of the Twelfth International Conference on. IEEE, 536–545.
14. Chu, C.-S. J., 1995. Time series segmentation: A sliding window approach. Inf. Sci. 85:147–173.
15. Hunter, J., and N. McIntosh, 1999. Knowledge-based event detection in complex time series data. In Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making. Springer, 271–280.
16. Vasko, K. T., and H. T. Toivonen, 2002. Estimating the number of segments in time series data using permutation tests. In Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on. IEEE, 466–473.
17. Terzi, E., and P. Tsaparas, 2006. Efficient algorithms for sequence segmentation. In Proceedings of the 2006 SIAM International Conference on Data Mining. SIAM, 316–327.
18. Little, M., and N. Jones, 2011. Generalized methods and solvers for noise removal from piecewise constant signals. I. Background theory. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 467:3088–3114.
19. Chen, J., and A. K. Gupta, 2001. On change point detection and estimation. Communications in statistics-simulation and computation 30:665–697.
20. Page, E. S., 1955. A Test for a Change in a Parameter Occurring at an Unknown Point. Biometrika 42:523–527.
21. Page, E. S., 1957. On Problems in which a Change in a Parameter Occurs at an Unknown Point. Biometrika 44:248–252.
22. Wiggins, P. A., 2015. An information-based approach to change-point analysis with applications to biophysics and cell biology. Biophys J 109:346–54.
23. LaMont, C. H., and P. A. Wiggins, 2016. The development of an information criterion for change-point analysis. Neural computation 28:594–612.
24. Watkins, L. P., and H. Yang, 2005. Detection of Intensity Change Points in Time-Resolved Single-Molecule Measurements. The Journal of Physical Chemistry B 109:617–628. PMID: 16851054.
25. Montiel, D., H. Cang, and H. Yang, 2006. Quantitative characterization of changes in dynamical behavior for single-particle tracking studies. The Journal of Physical Chemistry B 110:19763–19770.
26. Manzo, C., and L. Finzi, 2010. Quantitative analysis of DNA-looping kinetics from tethered particle motion experiments. Methods in enzymology 475:199–220.
27. Little, M., and N. Jones, 2013. Signal processing for molecular and cellular biological physics: an emerging field. Philosophical Transactions of the Royal Society of London A:

Mathematical, Physical and Engineering Sciences 371.
28. Duderstadt, K. E., H. J. Geertsema, S. A. Stratmann, C. M. Punter, A. W. Kulczyk, C. C. Richardson, and A. M. van Oijen, 2016. Simultaneous real-time imaging of leading and lagging strand synthesis reveals the coordination dynamics of single replisomes. Molecular cell 64:1035–1047.
29. Yin, S., N. Song, and H. Yang, 2017. Detection of Velocity and Diffusion Coefficient Change Points in Single-Particle Trajectories. Biophysical journal https://doi.org/10.1016/j.bpj.2017.11.008.
30. Chen, J., 1998. Testing for a change point in linear regression models. Communications in Statistics-Theory and Methods 27:2481–2493.
31. Nikol'skii, I., and K. Furmanov, 2016. Parallel algorithm to detect structural changes in time series. Computational Mathematics and Modeling 27:247–253.
32. Song, N., and H. Yang, 2017. Parallelization of change point detection. The Journal of Physical Chemistry A 121:5100–5109.
33. Kenney, J., and E. Keeping, 1947. Mathematics of statistics. Van Nostrand.
34. Singh, V., 2013. Entropy-Based Parameter Estimation in Hydrology. Water Science and Technology Library. Springer Netherlands.
35. Snedecor, G., and W. Cochran, 1991. Statistical Methods. Wiley.
36. Maheshwari, G., G. Brown, D. A. Lauffenburger, A. Wells, and L. G. Griffith, 2000. Cell adhesion and motility depend on nanoscale RGD clustering. J. Cell Sci. 113:1677–1686.
37. Bangasser, B. L., S. S. Rosenfeld, and D. J. Odde, 2013. Determinants of maximal force transmission in a motor-clutch model of cell traction in a compliant microenvironment. Biophys. J. 105:581–592.
38. Tempestini, A., C. Monico, L. Gardini, F. Vanzi, F. S. Pavone, and M. Capitanio, 2018. Sliding of a single lac repressor protein along DNA is tuned by DNA sequence and molecular switching. Nucleic Acids Research https://doi.org/10.1093/nar/gky208.
39. Ley, K., C. Laudanna, M. I. Cybulsky, and S. Nourshargh, 2007. Getting to the site of inflammation: the leukocyte adhesion cascade updated. Nat. Rev. Immunol. 7:678.
40. Makgoba, M. W., M. E. Sanders, G. E. G. Luce, M. L. Dustint, T. A. Springer, E. A. Clark, P. Mannoni, and S. Shaw, 1988. ICAM-1 a ligand for LFA-1-dependent adhesion of B, T and myeloid cells. Nature 331:86–88.
41. Oppenheimer-Marks, N., L. Davis, D. T. Bogue, J. Ramberg, and P. Lipsky, 1991. Differential utilization of ICAM-1 and VCAM-1 during the adhesion and transendothelial migration of human T lymphocytes. J. Immunol. 147:2913–2921.
42. Hood, J. D., and D. A. Cheresh, 2002. Role of integrins in cell invasion and migration. Nat. Rev. Cancer 2:91.
43. Kim, M., C. V. Carman, W. Yang, A. Salas, and T. A. Springer, 2004. The primacy of affinity over clustering in regulation of adhesiveness of the integrin αLβ2. J. Cell Biol. 167:1241–1253.
44. van Zanten, T. S., A. Cambi, M. Koopman, B. Joosten, C. G. Figdor, and M. F. Garcia-Parajo, 2009. Hotspots of GPI-anchored proteins and integrin nanoclusters function as nucleation sites for cell adhesion. Proc. Natl. Acad. Sci. U.S.A. 106:18557–18562.
45. Gelles, J., B. J. Schnapp, and M. P. Sheetz, 1988. Tracking kinesin-driven movements with nanometre-scale precision. Nature 331:450–453.
46. Blatz, A. L., and K. L. Magleby, 1986. Correcting single channel data for missed events. Biophys. J. 49:967–980.
47. Vanzi, F., C. Broggio, L. Sacconi, and F. S. Pavone, 2006. Lac repressor hinge flexibility and DNA looping: single molecule kinetics by tethered particle motion. Nucleic Acids Res. 34:3409–3420.
48. Capitanio, M., R. Cicchi, and F. Pavone, 2005. Position control and optical manipulation for nanotechnology applications. Eur. Phys. J. B 46:1–8.
49. Capitanio, M., D. Maggi, F. Vanzi, and F. Pavone, 2007. FIONA in the trap: the advantages of combining optical tweezers and fluorescence. J. Opt. A: Pure Appl. Opt. 9:S157.
50. Gittes, F., and C. F. Schmidt, 1998. Interference model for back-focal-plane displacement detection in optical tweezers. Opt. Lett. 23:7–9.
51. Capitanio, M., G. Romano, R. Ballerini, M. Giuntini, F. Pavone, D. Dunlap, and L. Finzi, 2002. Calibration of optical tweezers with differential interference contrast signals. Rev. Sci. Instrum. 73:1687–1696.
52. Capitanio, M., M. Canepari, P. Cacciafesta, V. Lombardi, R. Cicchi, M. Maffei, F. Pavone, and R. Bottinelli, 2006. Two independent mechanical events in the interaction cycle of skeletal muscle myosin with actin. Proc. Natl. Acad. Sci. U.S.A. 103:87–92.

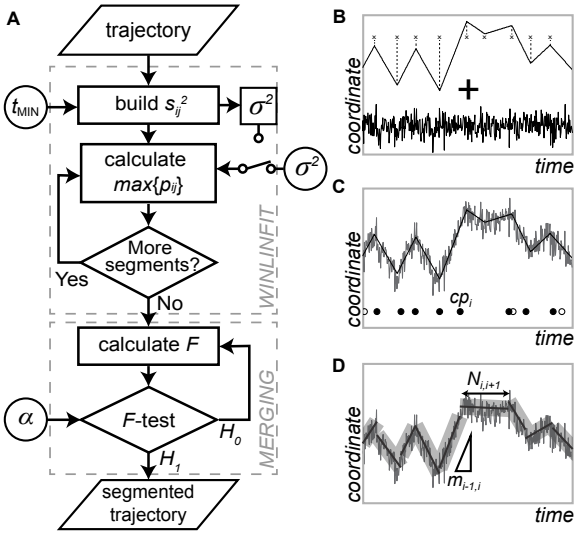

Figure 1: The principles of the PLANT algorithm. (A) Flow chart of the algorithm. (B) Representative simulated input trajectory as a series of joint linear segments corrupted with Gaussian noise. (C) The WINLINFIT block detects points compatible with a change of slope (full and empty symbols), which are further refined by the MERGING block based on their statistical significance (full symbols). (D) Reconstructed trajectory, from which it is possible to calculate the slopes and durations of each segment as well as their errors.

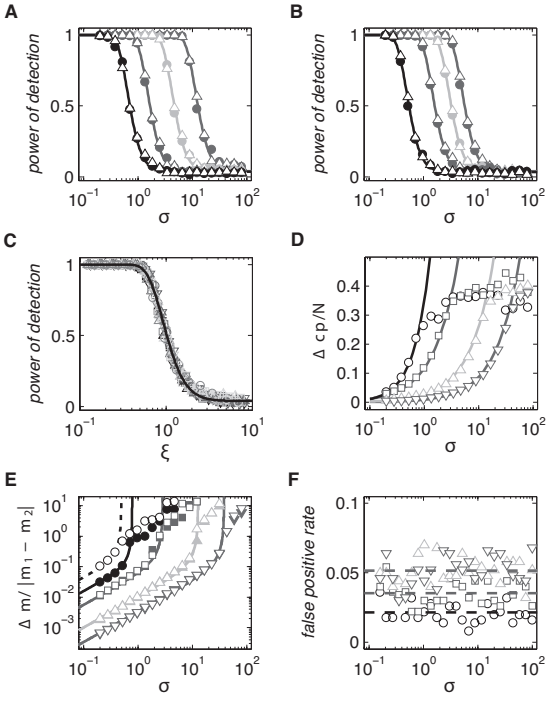

Figure 2: Algorithm performance for cp detection calculated on the simulation scheme depicted in Fig. S1A. (A) Power of detection as a function of the noise σ for segment lengths $N_1 = N_2 = 10, 20, 40, 80$ (from left to right, represented in different tones of gray) and slopes $m_1 = 0.2, 0.5$, $m_2 = 0.5, 0.8$ (● and △, respectively) with same slope increment $|m_2 - m_1| = 0.3$. (B) Power of detection as a function of the noise σ for segment slopes $m_1 = 0.2$ and $m_2 = 0.3, 0.5, 0.8, 1.3$ (from left to right, represented in different tones of gray) and lengths $N_1 = 13, N_2 = 30$ (●) and $N_1 = N_2 = 20$ (△) with similar N^* ($N^* \simeq 18$ and $N^* = 20$, respectively). (C) Power of detection for several segments lengths and slopes collapse onto each other once expressed in terms of the rescaled variable ξ, defined in Eq. S6. Lines in (A-C) correspond to the function given in Eq. S7. (D) Change point localization precision normalized to the trace length for traces composed by: segments with slopes $m_1 = 0.2, m_2 = 0.3$ and lengths $N_1 = 20$, $N_2 = 40$ (○); segments with slopes $m_1 = 0.2, m_2 = 0.3$ and lengths $N_1 = N_2 = 80$ (□); segments with slopes $m_1 = 0.2$, $m_2 = 0.5$ and lengths ($N_1 = N_2 = 80$) (△) ; segmants with slopes $m_1 = 0.2$, $m_2 = 1.1$ and lengths ($N_1 = N_2 = 80$) (▽). Lines correspond to Eq. S3. (E) Error in the determination of the segment slopes normalized to the modulus of slopes difference. Symbols have the same meaning as in panel D. Empty or filled symbols refer to the two different segments of the simulated trace. Lines correspond to Eq. S5. (F) False positive identification rate obtained for $\alpha = 0.05$ on traces of constant slope as a function of σ for different parameter sets: $m = 0.1, N = 40$ (○), $m = 0.1, N = 80$ (□), $m = 0.1, N = 150$ (△) and $m = 1.1, N = 150$ (▽). Each point in the plot was obtained from 500 simulated traces. Dashed lines correspond to the average values.

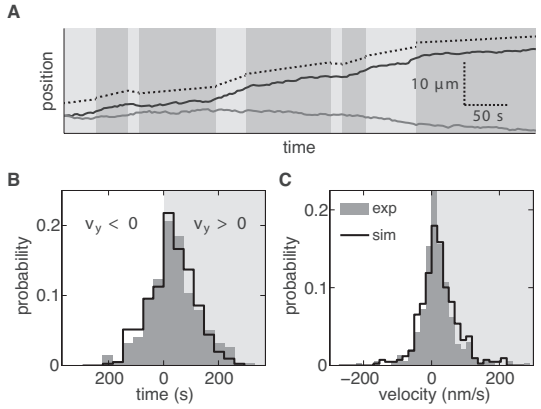

Figure 3: Analysis of a representative time traces from cell tracking experiments. (A) Time traces of the x (gray line) and y coordinates (black line) of the position of the cell represented in Fig. S3. Shaded gray regions correspond to regions where the algorithm detected changes of slope. The dashed line corresponds to the piecewise reconstructed trace of the trajectory along the y-direction that has been shifted upwards for clarity. (B) Histogram of the duration of the detected segments, divided according to the direction of the y-component of the velocity, along (light gray shaded area) or against (white area) the shear flow. (C) Histogram of the y-component of the velocity of the detected segments. In both panels, the gray bars correspond to experimental data, and the black lines are the resulting distributions of time and velocity after applying the algorithm to simulated data.

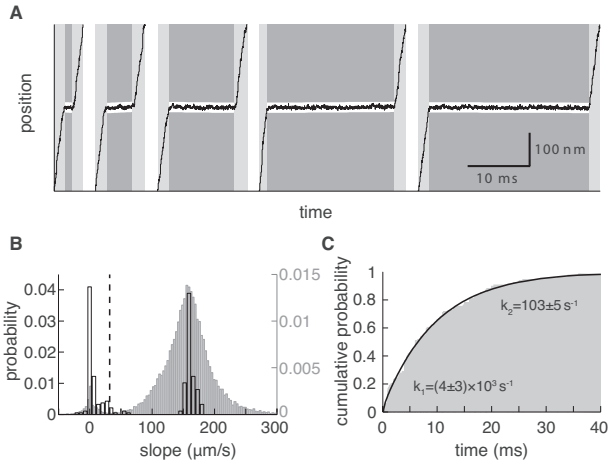

Figure 4: Analysis of ultrafast force-clamp spectroscopy experiments. (A) Fragments of traces corresponding to actin-myosin binding events with different durations under positive force of 1 pN and [ATP] =20 μM. Dark grey areas delimit the nearly flat regions found by the PLANT algorithm, corresponding to binding events. White lines represent the reconstructed segments. (B) Histograms of the velocity of the fragments before (gray bars) and after (open bars) threshold-based classification. The dashed black line corresponds to the threshold value. (C) Cumulative histograms of the duration of actin-myosin binding events (gray) and corresponding fit using a 2-exponential cumulative probability density function (black line).

Supporting Material

PLANT: A Method for Detecting Changes of Slope in Noisy Trajectories

Alberto Sosa-Costa[1], Izabela K. Piechocka[1], Lucia Gardini[2,3], Francesco S. Pavone[2,3,4], Marco Capitanio[2,4], Maria F. Garcia-Parajo[1,5], Carlo Manzo[1,6,*]

[1]ICFO-Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, Castelldefels, Barcelona, Spain; [2]LENS - European Laboratory for Non-linear Spectroscopy, Sesto Fiorentino, Italy; [3]National Institute of Optics–National Research Council, Florence, Italy; [4]Department of Physics and Astronomy, University of Florence, Sesto Fiorentino, Italy; [5]ICREA, Barcelona, Spain; and [6]Universitat de Vic - Universitat Central de Catalunya, Vic, Spain.

*Correspondence: carlo.manzo@uvic.cat

SUPPORTING MATERIALS AND METHODS

Automatic estimation of σ^2

When analyzing the time traces, our algorithm needs an estimation of the noise variance σ^2. In many cases, this value can be obtained from calibration measurements, e.g. by tracking immobile objects. However, we devised a method that allows for its online estimation directly from the trace under analysis. The working hypothesis consists in assuming that the time trace is the concatenation of segments of different slopes and duration. On such a trace, as shown in Eq. 4, we calculate the matrix s^2_{ij} for all the possible segments with length larger than N_{MIN}. The number of possible segments is given by Eq. 3. The s^2_{ij} calculated over regions fully contained in one of the segments of constant slope composing the trace represent a measurement of the variance σ^2, and their expectation value can be used to obtain an estimation of σ^2:

$$\sigma^2 = \frac{N}{N-1}\langle s^2 \rangle. \tag{S1}$$

However, not all the s^2_{ij} correspond to a measurement of the variance. Since we take all the possible segments, many will overlap with at least one border between two regions of different slope. However, for short segments (i.e., shorter than the average duration of the region of constant slopes), the number of times they lie at the border is negligibly small compared to the times they are fully contained within a constant slope region. By means of numerical simulations, we could show that in most experimental conditions the median of s^2_{ij} for segments with length ≤ 20 follows, save for a constant, Eq. S1. Therefore, we included in the algorithm the option of automatically estimating the noise variance by the weighted non-linear fitting of the median of the sample variance. The maximum segment length to be considered for the fit can be defined by the user to adapt to situations with a different average length for the regions of different slope.

Evaluation of the algorithm performance

The aim of the algorithm is to efficiently and precisely detect the occurrence of a change point corresponding to a variation of slope between two linear segments in the presence of noise. The trace noise, the slopes and the lengths of the adjacent segments affect the detection of the cp.In order to assess how these quantities influence the detection, we considered two simple simulation schemes (Fig. S1). The quantitation of the algorithm performance was obtained through the evaluation of several merit figures, such as the power of detection, the errors on the cp location and on the segment slopes, and the false positive rate. The behavior of these figures as a function of trace parameters was calculated according to the simulation scheme of (Fig. S1 A) and is discussed in the main manuscript (Fig. 2). However, geometrical and dimensional arguments can provide an estimation of the expected behavior. Let's first consider a change point cp separating two segments $i = 1, 2$ of slopes m_i and lengths N_i (Fig. S1 A). The trajectory coordinate is affected by Gaussian noise with standard deviation σ, therefore at each time point t its value shows deviations from the expected value $m_i t$. As a consequence, in a region around the change point position, we will find coordinate values compatibles with both slopes. To obtain a significative change (above a given confidence level) in the y coordinate that reveals the occurrence of the cp, one must be at a given distance Δ_{cp} from the cp, implying a minimum segment length that depends on the segment slope (Fig. S1 A). Following this semi-quantitative argument, we can obtain a simple geometrical estimation of how the cp localization precision depend on the trace parameters. In fact, by imposing the condition described above, i.e. $m_2\Delta_{cp} - \sigma \geq m_1\Delta_{cp} + \sigma$, we obtain:

$$\Delta_{cp} \approx \frac{2\sigma}{|m_2 - m_1|}. \tag{S2}$$

Although the simple calculation performed above predicts a dependence on noise and slopes only (Eq. S2), our simulations show a further effect of the segment length and a different scaling as a function of σ depending on the difference in segment length $\Delta N = |N_2 - N_1|$ (Fig. 2 D). We found that for small noise values the observed behavior is well described by the expression:

$$\Delta_{cp} \approx \frac{2\sqrt{2}N\sigma^{\frac{N}{N-\Delta N}}}{3N^* |m_2 - m_1|}, \tag{S3}$$

where $N = N_1 + N_2$ is the total trace length. For larger noise, the localization precision seems to asymptotically converge toward a value of $\sim 0.4N$, independently on the parameters values.

Following a similar reasoning as above, it is possible to estimate the dependence of the corresponding error on the slope. For a segment of length N_i, we can calculate the difference between the maximum and minimum slope compatible with the

extrema set by the finite cp localization and the noise σ, providing:

$$\Delta m_i \approx \frac{2\sigma\Delta_{cp}}{N_i^2 - \Delta_{cp}^2}. \tag{S4}$$

Also in this case, we found that the error on the detected slope shows a more complex dependence on the difference in segment lengths than the one reported in Eq. S4 that can be expressed for small noise as

$$\Delta m_i \approx \frac{4\sigma\sqrt{N_i}}{N_i^2 - \Delta_{cp}^2\left(\frac{N}{N+\Delta N}\right)^4}. \tag{S5}$$

The power of detection must be a function of all the three parameters characterizing the traces, i.e. segments slopes m_i, duration N_i and noise σ. We attempted to express this dependences in a more compact form based on dimensional considerations and numerical optimizations. We found that the power of detection for all of our simulated conditions collapsed on the same curve (Fig. 2 C) by expressing it in terms of the adimensional variable

$$\xi = \frac{2\sqrt{2\mathcal{N}}\sigma}{|m_2 - m_1|\, N^*\sqrt{N^*}}, \tag{S6}$$

where $\mathcal{N}$ is a constant with the dimensions of a segment length. Since the power of detection follows a sigmoidal behavior, empirically, we also found that it can be expressed as:

$$P_{cp}(\xi) = \frac{\mathrm{erf}\,(1/\xi)^4 + P_\infty}{1 + P_\infty}, \tag{S7}$$

where $\mathrm{erf}(\cdot)$ represents the error function, with $P_\infty \simeq 0.04$ and $\mathcal{N}^{1/2} \simeq 6$ for all the simulated conditions.

Although the scheme depicted in Fig. S1 A is useful to estimate power of detection and localization precision for an "isolated" cp, it represents a simplification of the actual situation, in which each segment of the trace (except the first and the last) is embedded within two others. In order to provide a flavour of the ability of the algorithm to correctly identify a segment in a more realistic scenario, we used another simulation scheme Fig. S1 B, where a test segment of varying slope and duration is embedded within two very long ones having null slope for different noise levels. The choice of the null slope does not limit the generality of these calculations since, as we have shown in Eq. S6 and Eq. S7, the power of detection only depends on the slope difference $|m_2 - m_1|$. For this simulation scheme, similarly to what we have previously discussed for the cp, a segment must verify some relation that depends on its length and slopes in order to be detected. As shown in Fig. S1 B for the case of outer segments with same slope equal to zero, this condition is given by $m_i \cdot N_i > 2\sigma$, which for a fixed slope, sets a minimum detectable-segment duration. However, it must be noticed that for large slopes this condition is verified even in the case of very short segments (e.g. $N_i \lesssim N_{min}$), for which the algorithm cannot detect two separated cp's but only one. However, since we do not force the reconstructed trace to be continuous at the cp, the trajectory will be reconstructed as having an instantaneous jump. In this way, in spite of missing the detection of the inner segment because of the limited time resolution, we could still preserve a high-fidelity reconstruction and the precision in the determination of the neighboring slopes.

The results of the algorithm applied over 500 simulations for each condition are reported in Fig. S2 B-C. Expectedly, the power of detection for a full segment has a nearly sigmoidal behavior (Fig. S2 B), similar to that reported for a single cp in Eq. S7. Deviations from this behavior can be observed for small segment length and slope, where the power of detection displays a little shoulder. This feature is likely due to the combination of the detection of a "real" cp plus a false positive detection, resulting in a spurious segment identification. In fact, this behavior is observed in the regime $N < 2\sigma/m$ (empty symbols in Fig. S2 B), where - as discussed previously - the values of noise, length and slope make the trace compatible with no segment change. As the simulated segment length and/or the slope are increased (or the noise reduced), the cp detection power increases and this also produces a decrease in the false positive rate. Successively, the power of detection for the full segment starts growing with the expected scaling. In order to show how the algorithm performs at different noise levels, we also report the dependence of $N_{50\%}$, i.e. the segment length at which the power of detection reaches 50%. According to Eq. S6, we find that indeed $N_{50\%}$ shows the expected power law behavior $\sim (\sigma/m)^{2/3}$ (Fig. S2 C).

Cell tracking experiments

The migration of immune cells across endothelial cells (ECs) during the immune response (1) involves interactions between the two cell types mainly mediated by the transient binding of adhesion receptors - called integrins - placed on the immune

cell membrane to their counter-receptors on the ECs surface (2, 3). The myriad of biochemical reactions occurring at the molecular level (4), influenced by molecular diffusion and by the spatial organization of cellular components, determine this complex cellular behavior. Moreover, it is known that, under certain conditions, integrins can spatially re-organize into clustered structures (5, 6). Such a clustering produces an increased local concentration that effectively changes the kinetics of interaction to their ligands (avidity). The kinetics of these interactions, in turn modulates cell migration on ECs (7). Thus, the changes in the motion of immune cells during the immune response are ultimately determined by the mechanism of interactions at molecular level. In addition, it must be considered that these phenomena take place in environments, such as lymph nodes and blood vessels, where mechanical cues such as shear forces play a major role. In order to study this mechanism in relation with the role of integrins in the immune function and the effect of mechanical forces, we performed cell tracking experiments under the presence of a laminar flow.

Cell culture and shear flow experiments

Human endothelial cell line (ECs) and Jurkat T lymphoblastoid (lymphocytes used as a T-cell model) were a gift from A. Cambi (Nijmegen, The Netherlands). In a typical experiment, ECs were plated on coverslips functionalized with fibronectin (20 μg/ml, Sigma-Aldrich, St Louis, MO) and grown to sub-confluence in RPMI-1640 medium (Lonza, Verviers, Belgium). To mimic inflammatory conditions, samples were previously stimulated with Tumor Necrosis Factor α (TNFα) cytokine (10 ng/ml, Sigma-Aldrich, St Louis, MO) for 20 hours. TNFα-activated ECs were pre-exposed to four hours of continuous shear flow of 8 dyn/cm^2, prior to T-cells adhesion. Next, Jurkat cells, adjusted to 3×10^6ml^{-1} prior to the experiments, were flown on top of ECs at 0.3 dyn/cm^2 for 3 minutes using RPMI-1640 perfusion medium. The flow was then stopped for 3 minutes to allow cells to accumulate. Next, the post-flow of 1 dyn/cm^2 was introduced and kept constant for 20 minutes. The entire period of T-cells perfusion was recorded by differential interference contrast microscopy (DIC) at 3 frame/s. T-cells motion was analyzed frame by frame from DIC images using a custom-written Matlab code. A cross-correlation function was used to estimate changes in the position of the centroid of a cell over time, similarly as described before (8). In total, 60 single-cell trajectories collected from 5 independent experiments were analysed.

Microfluidics device and live-cell microscopy

Shear flow experiments were performed with a parallel-plate closed-flow chamber (FCS2, Bioptechs Inc., Butler, PA). A 40 mm coverslip with a pre-formed endothelial layer was ensembled into the chamber and then mounted on the stage of an inverted microscope (Olympus IX71) equipped with a $20 \times /0.5$NA air objective and a CMOS video camera (Thorlabs). One end of the flow chamber was connected to an automated syringe pump (FCS2 Micro-Perfusion Pump) whereas the other end went to a medium reservoir. The microfluidic system was kept at 37° C by means of a temperature chamber embedding the microscope stage.

Simulations of cell tracking experiments

The histograms in Fig. 3 B-C, corresponding to the experimental data (blue bars) are characterized by exponential tails at long times and velocities. Although one would be tempted to extract the characteristic times and velocities by simply fitting these experimental distributions to single exponential functions, this straightforward approach would result in an overestimation of these parameters with respect to the real values. This is due to the finite time resolution of the experiments and the performance of the detection method, that set a limit to the duration of the shortest detectable event (the dead time). In addition, missed short events would in turn produce artificial long events, biasing estimation of characteristic parameters (9). Therefore, the quantitative analysis of our data requires the knowledge of algorithm performance, obtained e.g. from simulations, in order to be able to correct this kind of artefacts. However, when dealing with "real" data, the situation is often too complicated to attempt to perform this kind of correction. As an example, in our case the probability of detecting a segment strongly depends on the (*a priori* unknown) distribution of segment duration and slopes, as well as on the relative positioning of segments. Correcting the parameters estimation in an analytical fashion as done for other cases (9, 10) will in this case be a hard task. To circumvent this difficulty and obtain a robust estimation of the kinetic parameters, we performed numerical simulations.The results of these fits were compared to the results of the fits performed on the experimental data after being segmented. We created 50 synthetic traces with duration of 450 s composed by contiguous segments, with randomly generated duration and slope. The traces were corrupted with the noise value $\sigma = 0.2$ estimated for the actual data. We assumed exponential distributions for segment duration and slope similar to the experimental data, but where the characteristic values of these distributions

and the relative probability of switching direction were varied. Since our algorithm will equally affect the simulated trajectories and the experimental ones, we varied the parameters of our simulated trajectories so that after applying our algorithm to them we could retrieve similar distributions (quantified by the χ^2 value) as the ones obtained after applying the algorithm to the experimental data.

Force-clamp spectroscopy experiments

Forces play an important role in numerous biological processes. A clear example is muscle contraction, in which the motor protein myosin II pulls the actin filaments. During contraction, myosin II transiently binds to an actin filament and induces a unitary displacement (working stroke) of the filament; the myosin then detaches and a new cycle starts over. The kinetics of these actin-myosin interactions has been shown to depend on the mechanical load and ATP concentration. Recently, using ultrafast force-clamp spectroscopy it has been possible to apply constant loads to a single motor domain of myosin and thus measure the kinetics of the interaction with microsecond time resolution (11). In these experiments, an actin filament is stretched between two optically trapped beads, while a third bead, coated with myosin II, is fixed to the coverslip (Fig. S4). The stretched filament is approached to the myosin-coated bead to allow the two molecules to interact. A constant force is then applied to the dumbbell by exerting a stronger force (ΔF) on one of the two traps. To maintain the two molecules close to each other, once the dumbbell has moved more than 200nm, the net force on actin filament is reversed. When the actin and myosin II are not bound, the dumbbell moves with a constant velocity $v = \Delta F/\gamma$, where γ is the viscous drag coefficient of the dumbbell. In contrast, when an actin-myosin bond is formed, the net force applied on the dumbbell is transmitted to myosin and the dumbbell quickly stops.

Ultrafast force-clamp spectroscopy setup

The experimental setup is described in detail in Refs. (11). Briefly, the setup combines double optical tweezers with fluorescence microscopy. The apparatus was stabilized to less than 1 nm with both passive and active stabilization (12, 13). Double optical tweezers were obtained by splitting a single laser source (Nd:YAG 1064 nm) and focused on the sample plane by the objective. Each trap could be moved along the actin filament direction by AODs placed in a plane conjugate to the back focal plane of the objective. Position of the trapped beads with respect to the trap center x was measured using quadrant detector photodiodes (QDPs) placed in a plane conjugate to the back focal plane of the condenser (14). The force F applied on each bead was measured from the displacement of the bead from the trap center x and from a calibration of the trap stiffness k, as $F = -kx$. Before each experiment, k was calibrated over the entire range of trap positions used during the experiment, with a power spectrum method (15). Trap stiffnesses in the range of $0.03 - 0.14$ pN/nm were used in the experiments. A custom software written in Labview controlled the force-clamp system. Data were acquired at 200-kHz sample rate with a data acquisition and generation board (NI-PCI-7830R), provided with a field-programmable gate array. For each data point acquired, the variation of the force from the clamped value and the correction to be applied to the traps position were calculated onboard, using the field-programmable gate array and a proportional algorithm. Correction signals were sent to two custom-built direct digital synthesizers driving the AODs. The measured feedback delay was $\sim 8\mu$s.

Myosin preparation and experiments

Myosin was extracted from mice gastrocnemius muscles and purified as previously described (16). The S1 subfragment was obtained by proteolytic digestion with papain. G-actin was extracted from rabbit muscle, biotinylated, polymerized to form F-actin filaments and labeled with rhodamine phalloidin. Polystyrene fluorescent beads (510 nm diameter, FC03F, Bangs Laboratories) were coated with neutravidin (31000, Pierce) and used in trapping experiments. Flow cells were prepared as previously described (16).

SUPPORTING REFERENCES

1. Ley, K., C. Laudanna, M. I. Cybulsky, and S. Nourshargh, 2007. Getting to the site of inflammation: the leukocyte adhesion cascade updated. *Nat. Rev. Immunol.* 7:678.
2. Makgoba, M. W., M. E. Sanders, G. E. G. Luce, M. L. Dustint, T. A. Springer, E. A. Clark, P. Mannoni, and S. Shaw, 1988. ICAM-1 a ligand for LFA-1-dependent adhesion of B, T and myeloid cells. *Nature* 331:86–88.
3. Oppenheimer-Marks, N., L. Davis, D. T. Bogue, J. Ramberg, and P. Lipsky, 1991. Differential utilization of ICAM-1 and VCAM-1 during the adhesion and transendothelial migration of human T lymphocytes. *J. Immunol.* 147:2913–2921.

4. Hood, J. D., and D. A. Cheresh, 2002. Role of integrins in cell invasion and migration. *Nat. Rev. Cancer* 2:91.
5. Kim, M., C. V. Carman, W. Yang, A. Salas, and T. A. Springer, 2004. The primacy of affinity over clustering in regulation of adhesiveness of the integrin αLβ2. *J. Cell Biol.* 167:1241–1253.
6. van Zanten, T. S., A. Cambi, M. Koopman, B. Joosten, C. G. Figdor, and M. F. Garcia-Parajo, 2009. Hotspots of GPI-anchored proteins and integrin nanoclusters function as nucleation sites for cell adhesion. *Proc. Natl. Acad. Sci. U.S.A.* 106:18557–18562.
7. Maheshwari, G., G. Brown, D. A. Lauffenburger, A. Wells, and L. G. Griffith, 2000. Cell adhesion and motility depend on nanoscale RGD clustering. *J. Cell Sci.* 113:1677–1686.
8. Gelles, J., B. J. Schnapp, and M. P. Sheetz, 1988. Tracking kinesin-driven movements with nanometre-scale precision. *Nature* 331:450–453.
9. Blatz, A. L., and K. L. Magleby, 1986. Correcting single channel data for missed events. *Biophys. J.* 49:967–980.
10. Vanzi, F., C. Broggio, L. Sacconi, and F. S. Pavone, 2006. Lac repressor hinge flexibility and DNA looping: single molecule kinetics by tethered particle motion. *Nucleic Acids Res.* 34:3409–3420.
11. Capitanio, M., M. Canepari, M. Maffei, D. Beneventi, C. Monico, F. Vanzi, R. Bottinelli, and F. S. Pavone, 2012. Ultrafast force-clamp spectroscopy of single molecules reveals load dependence of myosin working stroke. *Nat. Methods* 9:1013–1019.
12. Capitanio, M., R. Cicchi, and F. Pavone, 2005. Position control and optical manipulation for nanotechnology applications. *Eur. Phys. J. B* 46:1–8.
13. Capitanio, M., D. Maggi, F. Vanzi, and F. Pavone, 2007. FIONA in the trap: the advantages of combining optical tweezers and fluorescence. *J. Opt. A: Pure Appl. Opt.* 9:S157.
14. Gittes, F., and C. F. Schmidt, 1998. Interference model for back-focal-plane displacement detection in optical tweezers. *Opt. Lett.* 23:7–9.
15. Capitanio, M., G. Romano, R. Ballerini, M. Giuntini, F. Pavone, D. Dunlap, and L. Finzi, 2002. Calibration of optical tweezers with differential interference contrast signals. *Rev. Sci. Instrum.* 73:1687–1696.
16. Capitanio, M., M. Canepari, P. Cacciafesta, V. Lombardi, R. Cicchi, M. Maffei, F. Pavone, and R. Bottinelli, 2006. Two independent mechanical events in the interaction cycle of skeletal muscle myosin with actin. *Proc. Natl. Acad. Sci. U.S.A.* 103:87–92.

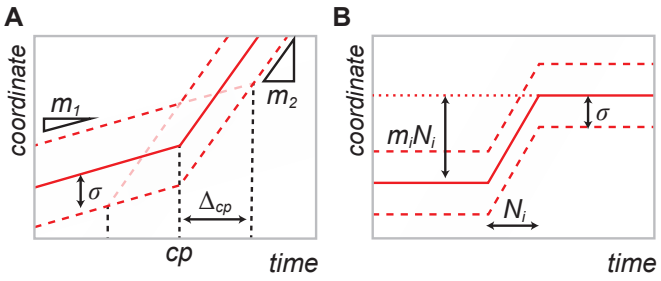

Figure S1: Simulations schemes used to calculate algorithm performance. (A) The power of detection of a cp was calculated by applying the algorithm on simulated traces formed by two segments with slopes m_1, m_2, lengths N_1, N_2 and noise σ; errors on the cp location and slopes were obtained from the same data. (B) The power of detection of a full segment was obtained by simulating segments of varying lengths N_i, slopes m_i and noise σ embedded within two much longer segments of null slope and same noise.

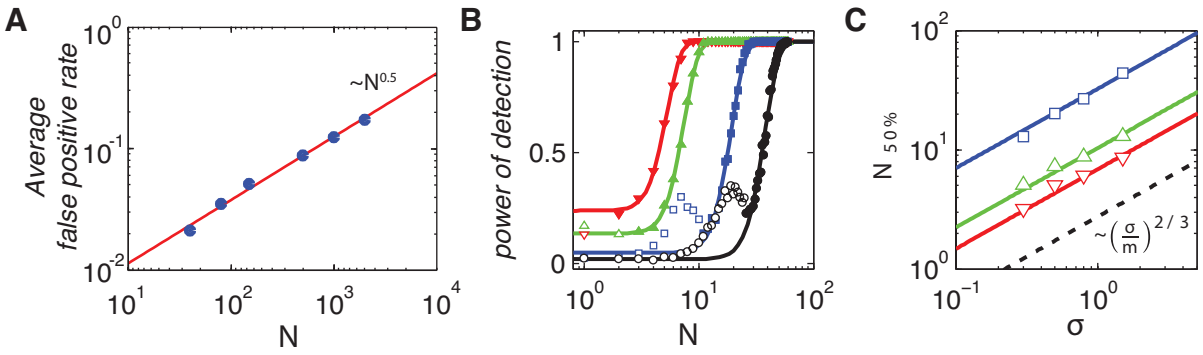

Figure S2: (A) Log-log plot of the average false positive rate (●) calculated according to the simulation scheme of Fig. S1 A as a function of the segment length N. The line is the result of a power law fit to the data, providing a scaling exponent ~ 0.5. (B) Power of detection according to the simulation scheme of Fig. S1 B as a function of the segment length N for $\sigma = 0.3$ and slopes m =0.04 (●), 0.1 (■), 0.5 (▲) and 1 (▼). The lines are the results of best fit according to Eq. S7. Empty symbols refer to the condition $N < 2\sigma/m$. (C) Segment length $N_{50\%}$ providing a 50% power of detection as a function of the trace noise σ for slopes m =0.1 (□), 0.5 (△) and 1 (▽). Lines correspond to a power law fit $\sim (\sigma/m)^{2/3}$. The dashed line is a guide to the eye.

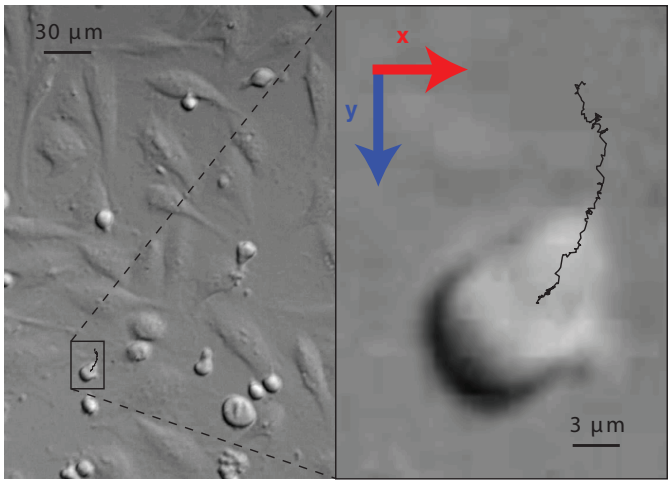

Figure S3: Cell tracking experiments. A frame of a typical experiment consisting in monitoring the motion of T-cells on ECs in the presence of shear flow (left) and zoom in of the region within the box (right) with overlaid trajectory. The flow is applied along to the positive y-direction.

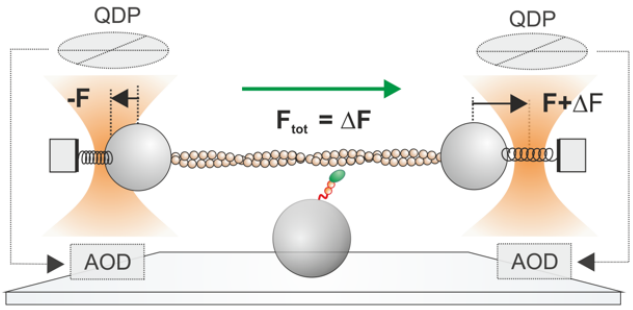

Figure S4: Schematic of the experimental setup for the ultrafast force-clamp spectroscopy experiments. An actin filament is stretched between two optically-trapped beads. A single myosin molecule is attached to a third bead stuck onto the cover-slip. Quadrant detector photodiodes (QDP) measure force, acousto-optic deflectors (AOD) rapidly stear the trapping beams to mantain the force on each bead constant. The net force applied to the actin filament (F_{tot}) is the sum of the forces clamped on each bead.