

Combination of density-clustering and supervised classification for event identification in single-molecule force spectroscopy data

Yongyi Yuan(袁泳怡), Jialun Liang(梁嘉伦), Chuang Tan(谭创),Xueying Yang(杨雪滢), Dongni Yang(杨东尼), and Jie Ma(马杰)

Citation: Chin. Phys. B, 2023, 32 (**10**): 108702. DOI: 10.1088/1674-1056/acf03e
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

New insight into the mechanism of DNA polymerase I revealed by single-molecule FRET studies of Klenow fragment

Rokshana Parvin, Qi Jia(贾棋), Jianbing Ma(马建兵), Chunhua Xu(徐春华), Ying Lu(陆颖), Fangfu Ye(叶方富), and Ming Li(李明)

Chin. Phys. B, 2022, 31 (**8**): 088701. DOI: 10.1088/1674-1056/ac6ed5

Dual mechanisms of Bcl-2 regulation in IP_3-receptor-mediated Ca^{2+} release: A computational study

Hong Qi(祁宏), Zhi-Qiang Shi(史志强), Zhi-Chao Li(李智超), Chang-Jun Sun(孙长君), Shi-Miao Wang(王世苗), Xiang Li(李翔), and Jian-Wei Shuai(帅建伟)

Chin. Phys. B, 2021, 30 (**10**): 108704. DOI: 10.1088/1674-1056/ac1e0d

Improved data analysis method of single-molecule experiments based on probability optimization

Weili Zhai(翟伟利), Guohua Yuan(袁国华), Chao Liu(刘超), Hu Chen(陈虎)

Chin. Phys. B, 2018, 27 (**1**): 018703. DOI: 10.1088/1674-1056/27/1/018703

Biexponential distribution of open times of a toy channel model

Xiang Li(李翔), Jing-Jing Zhong(钟金金), Xue-Juan Gao(高学娟), Yu-Ning Wu(吴宇宁), Jian-Wei Shuai(帅建伟), Hong Qi(祁宏)

Chin. Phys. B, 2017, 26 (**12**): 128703. DOI: 10.1088/1674-1056/26/12/128703

Experimental research on the feature of an x-ray Talbot-Lau interferometer versus tube accelerating voltage

Wang Sheng-Hao (王圣浩), Margie P. Olbinado, Atsushi Momose, Han Hua-Jie (韩华杰), Hu Ren-Fang (胡仁芳), Wang Zhi-Li (王志立), Gao Kun (高昆), Zhang Kai (张凯), Zhu Pei-Ping (朱佩平), Wu Zi-Yu (吴自玉)

Chin. Phys. B, 2015, 24 (**6**): 068703. DOI: 10.1088/1674-1056/24/6/068703

Combination of density-clustering and supervised classification for event identification in single-molecule force spectroscopy data

Yongyi Yuan(袁泳怡)[1,2], Jialun Liang(梁嘉伦)[1,2,†], Chuang Tan(谭创)[1,2], Xueying Yang(杨雪滢)[1,2], Dongni Yang(杨东尼)[1,2], and Jie Ma(马杰)[1,2,‡]

[1] *School of Physics, Sun Yat-sen University, Guangzhou* 510275, *China*
[2] *State Key Laboratory of Optoelectronic Materials and Technologies, Sun Yat-sen University, Guangzhou* 510006, *China*

(Received 1 July 2023; revised manuscript received 28 July 2023; accepted manuscript online 15 August 2023)

Single-molecule force spectroscopy (SMFS) measurements of the dynamics of biomolecules typically require identifying massive events and states from large data sets, such as extracting rupture forces from force-extension curves (FECs) in pulling experiments and identifying states from extension-time trajectories (ETTs) in force-clamp experiments. The former is often accomplished manually and hence is time-consuming and laborious while the latter is always impeded by the presence of baseline drift. In this study, we attempt to accurately and automatically identify the events and states from SMFS experiments with a machine learning approach, which combines clustering and classification for event identification of SMFS (ACCESS). As demonstrated by analysis of a series of data sets, ACCESS can extract the rupture forces from FECs containing multiple unfolding steps and classify the rupture forces into the corresponding conformational transitions. Moreover, ACCESS successfully identifies the unfolded and folded states even though the ETTs display severe nonmonotonic baseline drift. Besides, ACCESS is straightforward in use as it requires only three easy-to-interpret parameters. As such, we anticipate that ACCESS will be a useful, easy-to-implement and high-performance tool for event and state identification across a range of single-molecule experiments.

Keywords: single-molecule force spectroscopy, data analysis, density-based clustering, supervised classification

PACS: 87.80.Nj, 82.37.–j, 87.15.H–, 07.05.Kf **DOI:** 10.1088/1674-1056/acf03e

1. Introduction

Single-molecule force spectroscopy (SMFS)[1] based on optical tweezers (OTs), atomic force microscopy (AFM) or magnetic tweezers (MTs), has become a powerful method for revealing enormous insight into the dynamics of biomolecules, such as DNA/RNA structural transitions,[2–4] protein folding/unfolding dynamics,[5–8] receptor—ligand binding/unbinding processes[9,10] and DNA-protein interactions.[11,12] Generally, SMFS requires a good number of data sets to produce adequate sampling distributions of events and states in order to precisely extract the kinetic and thermodynamics information. Progress has been made in the improvement of both the throughput of SMFS experiments and the lifetime of each molecule through the upgrade of instruments[13–17] and sample preparation method,[18–20] making it possible to obtain large data sets through parallel or repeated measurements. Nevertheless, it is still a heavy workload to manually identify and locate the events of interest from the enormous measurement data, which is crucial for further analysis. Therefore, an appropriate approach allowing event identification from SMFS experiments in a reliable, automated and reproducible way is clearly desirable.

SMFS experiments generally adopt two types of assays, including the pulling assay and the force-clamp assay. In a pulling assay, conformational transitions such as the unfolding of protein or nucleic acid structures generate abrupt drops in force at the rupture process in the force-extension curves (FECs), as shown in Figs. 1(a) and 1(b). The distribution of rupture forces can yield the zero-force unfolding rate constant (k), the distance to the transition state ($x^{\ddagger}$) and the height of the energy barrier ($\Delta G^{\ddagger}$).[21–23] However, the accurate deduction of those three parameters requires a large number of unfolding events and precise extraction of the rupture force from FECs. To address this need, many algorithms[24–28] have been developed. Although established algorithms are capable of locating the rupture force, some algorithms require many parameters, making them hard to use. Besides most of algorithms cannot classify the rupture forces to the corresponding conformational transitions when the FEC reveals several transitions due to the multiple secondary structures of the molecule. Finally, to our knowledge, it is difficult to accurately extract the rupture forces when the molecule hops quickly between the unfolded state and the folded state during pulling.

In the second type of assay, i.e., the force-clamp assay, a constant force is applied to the molecule and the extension-time trajectories (ETTs) are recorded, where the events are indicated by the extension signal hopping between two states,

†Corresponding author. E-mail: liangjlun3@mail.sysu.edu.cn
‡Corresponding author. E-mail: majie6@mail.sysu.edu.cn

© 2023 Chinese Physical Society and IOP Publishing Ltd

as shown in Figs. 1(c) and 1(d). The distributions of the lifetimes of each state under different forces yield k, $x^{\ddagger}$ and $\Delta G^{\ddagger}$, and the free energy landscape can be reconstructed subsequently.[4,29] Therefore, it is critical to identify each state in ETTs. Hidden Markov model (HMM)[30–32] is widely adopted in SMFS and single-molecule fluorescence resonance energy transfer (smFRET) experiments for state detection. Recently, algorithms based on machine learning such as CPHAC[33] and DISC[34] have been developed to identify multiple states in single molecule trajectories with high accuracy and high computational performance. However, all these algorithms require the input data to be free of base-line drift. A reference-bead strategy is widely used in MTs experiments for drift correction,[7,35] but this strategy may not be suitable for other single molecule experiments and also probably cannot completely correct the drift of all tethered beads in high-throughput experiments. Although other progress has been made in drift correction in OT experiments, it has to be compensated with the increase of either equipment cost or experiment complexity.[36–39]

Here, we present a solution, named approach combining clustering and classification for event identification of SMFS (ACCESS), for locating and identifying both the rupture forces in FECs and the transition states in ETTs. In pulling experiments, ACCESS is capable of segmenting the FEC, locating the rupture force and classifying the rupture force to the corresponding transitions. In force-clamp experiments, ACCESS can accurately identify the folded and unfolded states for ETTs with severe nonmonotonic drift. In addition, ACCESS is straightforward to implement and easy to understand because it requires just three easy-to-interpret parameters. We demonstrate the performance of ACCESS for three diverse data sets, including simulated FECs, experimentally measured FECs from OTs, and ETTs from MTs, and show its potential in improving the fully automated SMFS analysis.

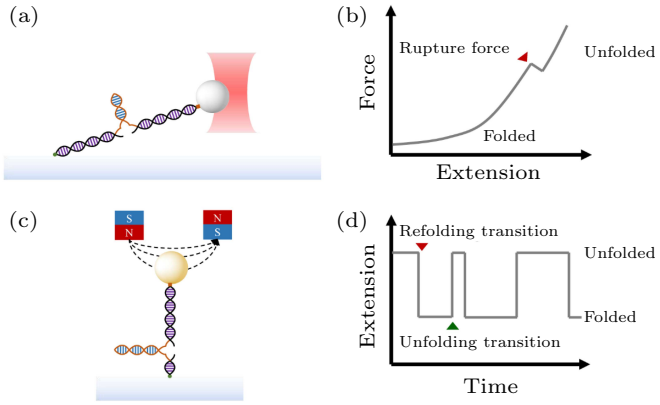


Fig. 1. Schematic of SMFS experiments. (a) Schematic of DNA pulling experiments with a single-trap optical tweezer. (b) Schematic diagram of a force-extension curve. The DNA hairpin switches from folded state F to unfolded state U at rupture force. (c) Schematic of DNA hopping experiments using a magnetic tweezer. (d) Schematic diagram of an extension-time trajectory. Near the critical force, DNA hairpin hops between the folded and unfolded states.

2. Materials and methods

ACCESS was written in MATLAB 2023a (the MathWorks, Natick, MA). ACCESS comprised of three major processes: clustering based on density-based spatial clustering of applications with noise (DBSCAN),[40–42] training and classification based on supervised learning.[43] The basis for the statistical analysis used by ACCESS and all performance metrics were presented in supplementary information (SI). To evaluate the performance of ACCESS, we used the following data sets: (i) 500 simulated force-extension curves (FECs) calculated from a free energy profile with folded (F), intermediate (I) and unfolded (U) states (Fig. S1). The simulated FECs revealed two conformational transitions, i.e., F $\rightarrow$ I and I $\rightarrow$ U, and several hopping between F (I) and I (U). (ii) 63 experimental FECs of DNA hairpin with 15 bp (base pairs) stem and 4 nt (nucleotides) loop measured by our home-built OTs in pulling experiments. Experimental FECs displayed WLC behavior of DNA handles at low force followed by hairpin unfolding at $\sim$ 13 pN. (iii) 7 experimental extension-time trajectories of DNA hairpin with 40 bp stem and 5 nt loop measured by our home-built MTs in force-clamp experiments. ETTs revealed two-state hopping behaviors with severe baseline drift. The FECs simulation and experimental details were presented in supplementary information. All analysis results were obtained using a desktop personal computer with 16 GB of RAM, Intel(R) Core (TM) i7-4790 CPU@ 3.60 Hz and a 305-GB hard drive.

3. Result and discussion

3.1. Extraction of rupture force in FECs using ACCESS

We first demonstrated the performance of ACCESS using simulated FECs. The workflow of ACCESS was shown in Fig. 2(a), which basically consisted of three steps, i.e., clustering, training, and classification. Given the input data, as shown in Fig. 2(b), ACCESS first clustered all data points of the FEC using the DBSCAN algorithm with two user-definable parameters, ε (the maximum radius of a neighborhood from the observing point) and MinPts (the minimum number of data points contained in a neighborhood). The settings of ε and MinPts were discussed in detail in (SI). In DBSCAN processing, all data points were preliminarily clustered into several groups, indicated as different colors and labeled numbers in Fig. 2(c). The data points labeled as -1 by DBSCAN were considered as outliers, indicated as the red crosses in Fig. 2(c). After excluding the outliers, the mean values of extension of each group were further clustered using the agglomerative hierarchical clustering (AHC) algorithm which generated a dendrogram as shown in Fig. S2. The dendrogram was cut into N_S groups, the third user-definable parameter, with the prior knowledge of the number of conformational states revealed by

FECs. For example, N_S was specified to three according to Fig. 2(b). As a result, groups in Fig. 2(c) with similar mean extension were combined together and finally the FEC was re-clustered into three classes with data points labeled with F, I, and U, as shown in Fig. 2(d). It should be noted that the settings of ε and MinPts were not unique. Different settings of ε and MinPts were also able to produce correct cluster results for the following analysis, as shown in Fig. S3. In step two, ACCESS generated a training model by utilizing the labeled FEC (Fig. 2(d)) as training data through an optimized support vector machine (SVM) algorithm since data points in FEC were linearly separable. In step three, ACCESS applied the training model to this FEC to capture transient transitions which were always identified as outliers in the prior steps, as described later. Moreover, since the feature of FECs was similar from trace to trace, the training model generated from a typical FEC could be applied to other FECs for classification, which significantly reduced the computational time. Figure 2(e) displayed the classification result of another simulated FEC by using the training model generated from Fig. 2(d). More results can be found in Fig. S4. The force was extracted as rupture forces[44] once the labels of data points changed from F to I or I to U. Therefore, ACCESS can classify the extracted rupture forces to the corresponding conformation changes. After analysis of 500 simulated FECs by ACCESS, Fig. 2(f) showed the distributions of the rupture forces for two types of transitions, F $\rightarrow$ I (blue histogram) and I $\rightarrow$ U (orange histogram). To verify the validity of rupture forces extracted by ACCESS, we fitted the two distributions using Eq. (S3), giving the fitting results of $\Delta G^{\ddagger}_{\mathrm{FI}} = 12.0 k_{\mathrm{B}}T$, $k^{0}_{\mathrm{FI}} = 0.0433\ \mathrm{s}^{-1}$, $x^{\ddagger}_{\mathrm{FI}} = 2.45$ nm (blue line) and $\Delta G^{\ddagger}_{\mathrm{IU}} = 22.0 k_{\mathrm{B}}T$, $k^{0}_{\mathrm{IU}} = 0.0050\ \mathrm{s}^{-1}$, $x^{\ddagger}_{\mathrm{IU}} = 2.24$ nm (orange line), which were consistent with the input parameters used in the FEC simulation (see Fig. S1). Therefore, ACCESS showed a good performance in locating and classifying the rupture forces.

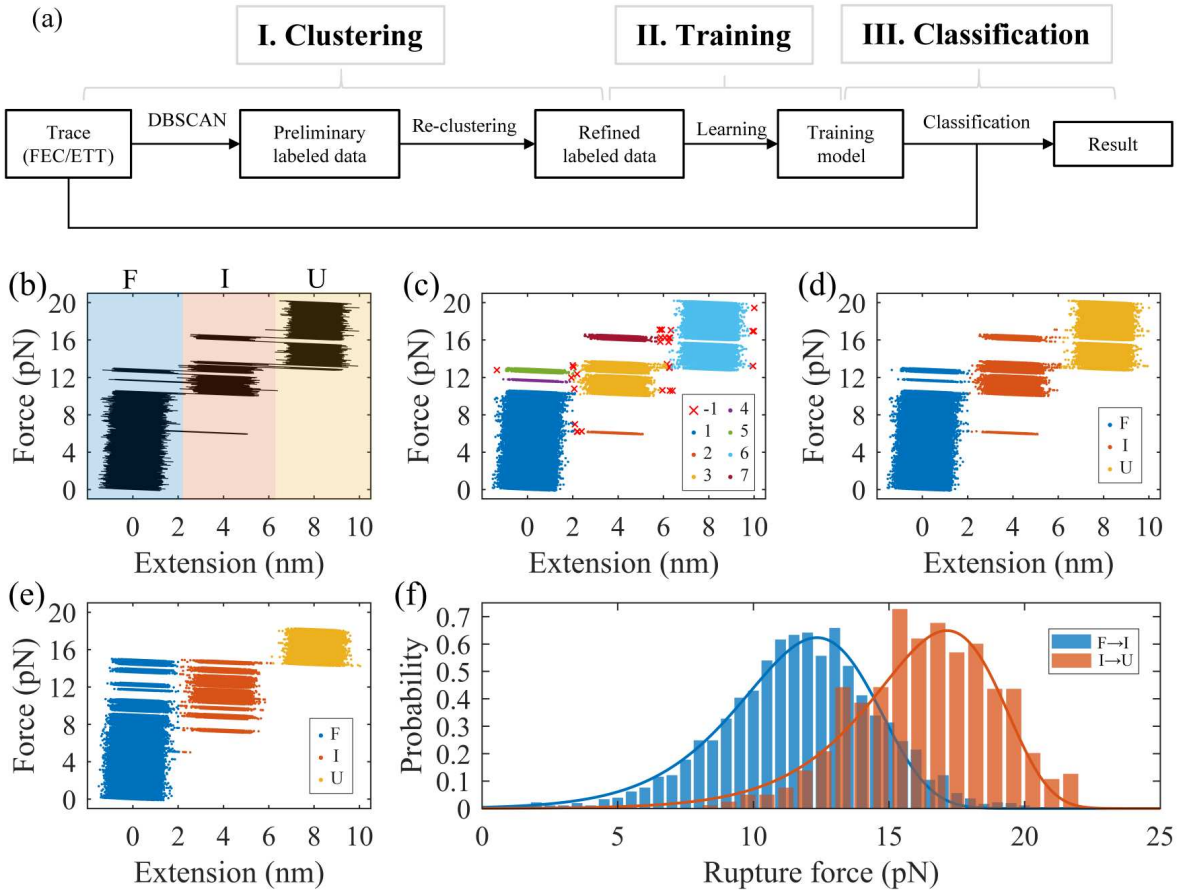


Fig. 2. An overview of analyzing simulated FECs using ACCESS. (a) The workflow of ACCESS for FEC/ETT analysis. The major steps of ACCESS were clustering, training and classification. (b) A typical simulated FEC reveals three conformational states, F, I and U. The FEC displayed several times of hopping between adjacent states. (c) The preliminary cluster result using DBSCAN. Data points of panel (b) were clustered into 7 groups as indicated by different labeled numbers and colors. The data points labeled as -1 by DBSCAN were identified as outliers. The parameter of DBSCAN was set as follows: $\varepsilon = 0.1$, MinPts $= 30$. More cluster results using different settings of ε and MinPts were found in Fig. S3. (d) The re-cluster result using AHC. The labeled data points of panel (c) were re-clustered into 3 classes, shown as F, I and U. (e) Classification result of another simulated FEC using training model generated from panel (d). (f) Distributions of rupture forces extracted from 500 simulated FECs by ACCESS. The blue line and red line were fittings using Eq. (S3) to the distributions of F $\rightarrow$ I transition (blue histogram) and I $\rightarrow$ U transition (orange histogram), respectively.

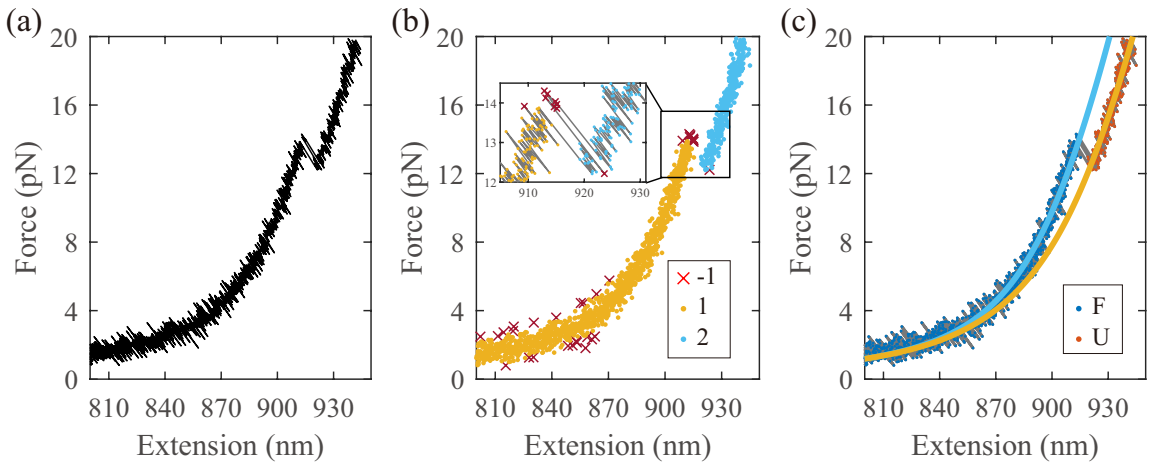


Fig. 3. Classification results of ACCESS on experimental FECs. (a) A typical experimental FEC of DNA hairpin with 2.6 kbp handles. (b) The preliminary cluster result of panel (a) using DBSCAN. DBSCAN clustered data points into two groups and identified outliers as −1. Inset: detail of the FEC showing that DNA hairpin hopped between the folded and unfolded states at ∼ 13.5 pN. Data points in transient transition were identified as outliers (red cross). The parameter of DBSCAN was set as follows: $\varepsilon = 0.055$, MinPts = 18. More cluster results using different settings of ε and MinPts were found in Fig. S6. (c) The classification result of panel (a) using the training model learning from panel (b). The data points identified as outliers in panel (b) were correctly classified into the corresponding state. WLC fitting (lines) to the two segments indicated the DNA hairpin of 33 nt.

We further applied ACCESS to analyze experimental FECs of a DNA hairpin. Figure 3(a) shows a typical FEC obtained by our home-built optical tweezers. Similarly, the FEC was firstly clustered into two groups with the outliers labeled as −1 by ACCESS, as shown in Fig. 3(b). The training model was generated from labeled data points without outliers and subsequently applied to all experimental FECs for data point classification. As the inset in Fig. 3(b) showed, data points in transient transitions were incorrectly identified as outliers by DBSCAN. In contrast, those data points were correctly classified into the corresponding states after being reclassified by the training model, as shown in Fig. 3(c). Moreover, the segmentation of the FEC was also convenient and useful for worm-like chain (WLC) fitting.[45] The number of nucleotides of DNA hairpin deduced from WLC fittings (blue line and yellow line in Fig. 3(c)) was 33 nt, which was close to the true value, i.e., 34 nt. More classification results of experimental FECs by using the training model generated from Fig. 3(b) could be found in Fig. S5.

3.2. Identification of transition states in ETTs using ACCESS

In this section, we illustrated how ACCESS could be applied to experimental ETTs with drift to identify the folded and unfolded transition states. Figure 4(a) showed an ETT of DNA hairpin measured at 15.5 pN by our home-built magnetic tweezers without using the reference bead for base-line correction. As a result, the ETT displayed significant nonmonotonic drift and could not be accurately analyzed by HMM (green line). The workflow of ACCESS for ETTs analysis was similar to Fig. 4(a). Briefly, ACCESS clustered the data points into several groups based on DBSCAN, as Fig. 4(b) showed. After eliminating the outliers, all groups were clustered into two classes, i.e., F and U, using a separate algorithm called label-change (details see supplementary information). Because data points in ETT with drift were linearly inseparable, ACCESS generated a training model by utilizing the labeled ETT (Fig. 4(c)) as training data through an optimized neural network (NN) algorithm[46] instead of SVM. The training model was applied to this ETT for reclassification. As mentioned above, the data points in transient transitions were always identified as outliers by DBSCAN at first, indicated as the red crosses in the inset in Fig. 4(b) and could be classified into the correct states after reclassification by the training model. The final result is shown in Fig. 4(d). As can be seen, the unfolded state and folded state were accurately identified by ACCESS, even for those in the transient transitions. It should be emphasized that, since the pattern of drift was random and different from trace to trace, the training model generated from one trace was unable to correctly classify other traces. Therefore, each drifted ETT needed to generate its own training model for classification.

To prove the validity of ACCESS on drifted ETTs with different kinetic hopping rates, we further applied ACCESS to the identification of conformational states of a DNA hairpin clamped at different forces. As Fig. 5 showed, a DNA hairpin was clamped at 16.0 pN, 15.8 pN, 15.7 pN, 15.6 pN, 15.4 pN, 15.3 pN from panel (a) to panel (f), respectively, exhibiting different lifetime of the folded and unfolded states. As expected, ACCESS was able to precisely identify the two states at different forces. These results allowed further kinetic analysis and reconstruction of the free energy landscape.

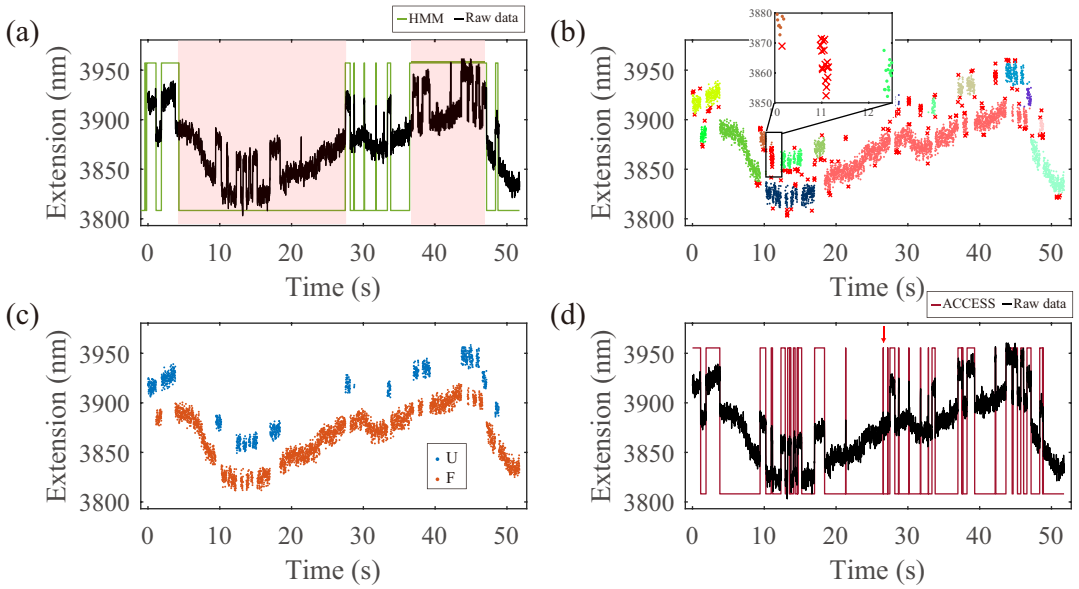


Fig. 4. An overview of analyzing experimental ETTs using ACCESS. (a) An extension-time trajectory of DNA hopping at $F = 15.5$ pN with baseline drift. The green line is the analysis result by HMM. The red background indicates the misidentification of states. (b) The preliminary cluster result of ETT (a) using DBSCAN. DBSCAN clustered data points of panel (b) into 15 groups and identified outliers as red crosses. Here, ε and MinPts were set as 0.08 and 12, respectively. More cluster results using different settings of ε and MinPts were found in Fig. S7. Inset: detail of the ETT showing that data points implying the short lifetime of the state were identified as outliers. (c) The data points of panel (b) were further divided into two classes by the label-change algorithm. (d) The classification result of panel (a) using the training model learning from panel (c). The red arrow represents false negative indicating misidentified events.

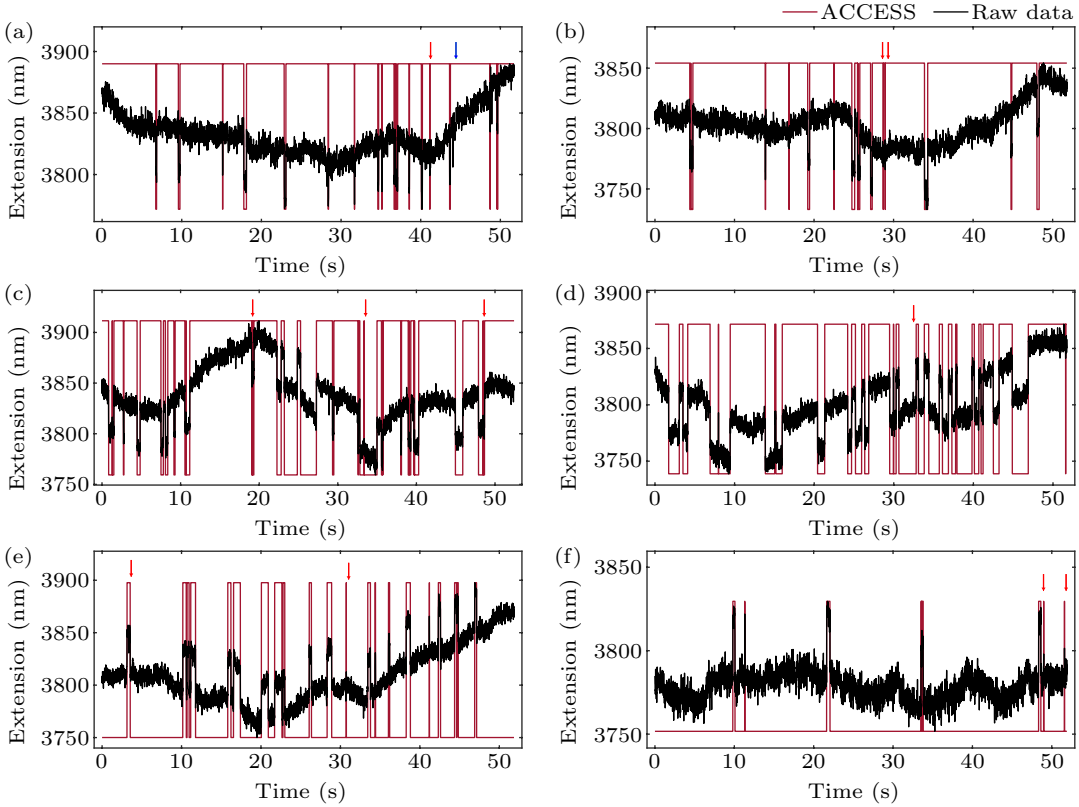


Fig. 5. The analysis result of experimental ETTs using ACCESS. The DNA hairpin with 12 kbp DNA handles is clamped at 16.0 pN, 15.8 pN, 15.7 pN, 15.6 pN, 15.4 pN and 15.3 pN from panel (a) to panel (f), respectively. The red arrow represents false negative indicating misidentified events, and the blue arrow represents false positive indicating missed events.

3.3. Evaluation of ACCESS's performance

Across all the applications, ACCESS provides high accuracy, high precision, and high recall (Table 1, see supplementary information for the definition of the performance metrics), inferred from the confusion matrix shown in Fig. S8. For extracting rupture forces from simulated FECs and experimental FECs, ACCESS performs exceptionally well with accuracy more than 95%. Not only does ACCESS provide high accu-

racy, it is also fast. For example, the total analysis time of 64 experimental FECs with $\sim$ 4000 data points per trajectory is $\sim$ 6 s. For force-clamp experiments, ACCESS can return a high accuracy of 94.49% for experimental ETTs with large drift while for this case HMM method failed in correctly recognizing most of the folded and unfolded states. The small error left may be caused by the noise points, as shown in Fig. 4. The analysis time of an ETT with $\sim$ 5000 data points is $\sim$ 17 s by ACCESS. It should be noted that all the data sets input into ACCESS are not smoothed. We believe that the performance metrics will be further improved after data smoothing.

Table 1. Evaluation results for ACCESS on different data sets.

	Accuracy	Precision	Recall
Simulated FEC	98.78%	100%	98.78%
Experimental FEC	96.67%	98.31%	98.31%
Experimental ETT	94.49%	99.55%	94.89%

4. Conclusion

We demonstrate the capability of ACCESS, based on clustering and supervised learning, to automatically analyze SMFS data. In pulling experiments, ACCESS can segment the FECs, extract the rupture forces, and classify the rupture forces into the corresponding conformational transitions, as demonstrated by the analysis of simulated FECs and experimental FECs. Additionally, ACCESS can analyze FECs containing more than two transitions. However, ACCESS cannot provide correct results if the data sets contain too many meaningless and uninterpretable spectra. Therefore, those spectra should be ruled out first before performing ACCESS analysis. Moreover, in force-clamp experiments, unlike previous approaches that require the input data to be free of base-line drift, ACCESS is able to identify the folded and unfolded states even though the ETTs suffer from severe nonmonotonic drifting, which is essential for further kinetic analysis. When dealing with extremely large data sets, such as ETTs collected in serval hours, it will take a lot of time to generate a training model from the ETTs. One way to solve this problem is to cut the long-time ETTs into several segments and individually analyze each segment. ACCESS requires only three easy-to-interpret parameters, ε and MinPts for DBSCAN and N_S for re-clustering. Combining DBSCAN and supervised classification, ACCESS efficiently identifies the fast transitions in SMFS data. Its high accuracy and fast execution allow for the automatic analysis of large data sets, reducing human labor and intervention. In addition to applications in optical tweezers and magnetic tweezers as demonstrated in this paper, ACCESS has great potential to be applied to other single molecule techniques, such as AFM and smFRET. It should be noted that the parameters, ε and MinPts for DBSCAN and N_S for re-clustering should be appropriately chosen since the data sampling rate and resolution may vary for different setups. Overall, we anticipate that ACCESS can be a useful, easy-to-implement and high-performance tool for single-molecule data analysis in a reliable, automated and reproducible way.

Acknowledgments

We thank members of the J. M. laboratory for critical discussion and comments and the support from the Physical Research Platform in the School of Physics of Sun Yat-sen University (PRPSP, SYSU). Project supported by the National Natural Science Foundation of China (Grant No. 12074445) and the Open Fund of the State Key Laboratory of Optoelectronic Materials and Technologies of Sun Yat-sen University (Grant No. OEMT-2022-ZTS-05).

References

[1] Neuman K C and Nagy A 2008 *Nat. Methods* **5** 491
[2] Woodside M T and Block S M 2014 *Annu. Rev. Biophys.* **43** 19
[3] Neupane K, Zhao M, Lyons A, Munshi S, Ileperuma S M, Ritchie D B, Hoffer N Q, Narayan A and Woodside M T 2021 *Nat. Commun.* **12** 4749
[4] Woodside M T, Behnke-Parks W M, Larizadeh K, Travers K, Herschlag D and Block S M 2006 *Proc. Natl. Acad. Sci. USA* **103** 6190
[5] Rief M, Gautel M, Oesterhelt F, Fernandez J M and Gaub H E 1997 *Science* **276** 1109
[6] Bustamante C, Alexander L, Maciuba K and Kaiser C M 2020 *Annu. Rev. Biochem.* **89** 443
[7] Guo Z, Hong H, Yuan G, Qian H, Li B, Cao Y, Wang W, Wu C X and Chen H 2020 *Phys. Rev. Lett.* **125** 198101
[8] Lei H, Zhang J, Li Y, Wang X, Qin M, Wang W and Cao Y 2022 *ACS Nano* **16** 15440
[9] Merkel R, Nassoy P, Leung A, Ritchie K and Evans E 1999 *Nature* **397** 50
[10] Ott W, Jobst M A, Schoeler C, Gaub H E and Nash M A 2017 *J. Struct. Biol.* **197** 3
[11] Synakewicz M, Bauer D, Rief M and Itzhaki L S 2019 *Sci. Rep.* **9** 13820
[12] Liang L, Ma K, Wang Z, Janissen R and Yu Z 2021 *Biophys. J.* **120** 3283
[13] Cnossen J P, Dulin D and Dekker N H 2014 *Rev. Sci. Instrum.* **85** 103712
[14] Sitters G, Kamsma D, Thalhammer G, Ritsch-Marte M, Peterman E J G and Wuite G J L 2015 *Nat. Methods* **12** 47
[15] Agarwal R and Duderstadt K E 2020 *Nat. Commun.* **11** 4714
[16] Akbari E, Shahhosseini M, Robbins A, Poirier M G, Song J W and Castro C E 2022 *Nat. Commun.* **13** 6800
[17] Janissen R, Eslami-Mossallam B, Artsimovitch I, Depken M and Dekker N H 2022 *Cell Rep.* **39** 110749
[18] de Vlaminck I, Henighan T, Van Loenhout M, Pfeiffer A, Huijts J, Kerssemakers J W, Katan A, van Langen-Suurling A, van der Drift E, Wyman C and Dekker C 2011 *Nano Lett.* **11** 5489
[19] Janissen R, Berghuis B A, Dulin D, Wink M, van Laar T and Dekker N H 2014 *Nucleic Acids Res.* **42** e137
[20] Popa I, Rivas-Pardo J A, Eckels E C, Echelman D J, Badilla C L, Valle-Orero J and Fernández J M 2016 *J. Am. Chem. Soc.* **138** 10546
[21] Garai A, Zhang Y and Dudko O K 2014 *J. Chem. Phys.* **140** 135101
[22] Dudko O K, Hummer G and Szabo A 2006 *Phys. Rev. Lett.* **96** 108101
[23] Dudko O K, Hummer G and Szabo A 2008 *Proc. Natl. Acad. Sci. USA* **105** 15755
[24] Lin Z, Gao X, Li S and Hu C 2021 *Biochem. Biophys. Res. Commun.* **556** 59
[25] Partola K R and Lykotrafitis G 2016 *J. Biomech.* **49** 1221
[26] Sandal M, Benedetti F, Brucale M, Gomez-Casado A and Samorì B 2009 *Bioinformatics* **25** 1428
[27] Heenan P R and Perkins T T 2018 *Biophys. J.* **115** 757
[28] Duanis-Assaf T, Razvag Y and Reches M 2019 *Anal. Methods* **11** 4709

[29] Woodside M T, Anthony P C, Behnke-Parks W M, Larizadeh K, Herschlag D and Block S M 2006 *Science* **314** 1001
[30] McKinney S A, Joo C and Ha T 2006 *Biophys. J.* **91** 1941
[31] van de Meent J W, Bronson J E, Wiggins C H and Gonzalez R L J 2014 *Biophys. J.* **106** 1327
[32] Bronson J E, Fei J, Hofman J M, Gonzalez R L J and Wiggins C H 2009 *Biophys. J.* **97** 3196
[33] Hadzic M C A S, Börner R, König S L B, Kowerko D and Sigel R K O 2018 *J. Phys. Chem. B* **122** 6134
[34] White D S, Goldschen-Ohm M P, Goldsmith R H and Chanda B 2020 *eLife* **9** e53357
[35] Berghuis B A, Köber M, van Laar T and Dekker N H 2016 *Methods* **105** 90
[36] Carter A R, Seol Y and Perkins T T 2009 *Biophys. J.* **96** 2926
[37] Perkins T T 2014 *Annu. Rev. Biophys.* **43** 279
[38] Carter A R, King G M, Ulrich T A, Halsey W, Alchenberger D and Perkins T T 2007 *Appl. Opt.* **46** 421
[39] Nugent-Glandorf L and Perkins T T 2004 *Opt. Lett.* **29** 2611
[40] Lv Y, Ma T, Tang M, Cao J, Tian Y, Al-Dhelaan A and Al-Rodhaan M 2016 *Neurocomputing* **171** 9
[41] Jain A K, Murty M N and Flynn P J 1999 *ACM Comput. Surv.* **31** 264
[42] Sander J, Ester M, Kriegel H P and Xu X 1998 *Data Min. Knowl. Discov.* **2** 169
[43] Akinsola J E T 2017 *Int. J. Comput. Trends Technol.* **48** 128
[44] Zhang Y and Dudko O K 2013 *Proc. Natl. Acad. Sci. USA* **110** 16432
[45] Smith S B, Cui Y and Bustamante C 1996 *Science* **271** 795
[46] Sahay T, Aggarwal A, Bansal A and Chandra M 2015 2015 1*st Int. Conf. Gener. Comput. Technol. NGCT*, September 4–5, 2015, Dehradun, India, pp. 960–964