

Signal Recognition and Processing of Feature Information of Electrical Pulses of Nanopore Sensors *

ZENG Xiangjie, WANG Runyu, LUO Xingyue, WU Lingzhi*, WENG Lixing

(School of Geographic and Biological Information, Nanjing University of Posts and Telecommunications, Nanjing Jiangsu 210023, China)

Abstract: With the rapid development of nanopore sensors, growing demand appears for recognition and information extraction of nanopore electrical signals of various target molecules. How to process the weak electrical signals in the background of strong noise is a challenge for nanopore device. Thus, a signal processing program based on MATLAB has been designed to identify and extract feature information of electrical signals of nanopore sensors for high-through detection. Firstly, the pulse signal of the selected current segment is recognized through the shifting of the analytical window, and fitted by using correction functions to reduce the interference of background noise. Once the characteristic event is identified, more information such as dwell time, current amplitude and peak area will be exported and saved. The designed program is helpful to the development of the information omics and precision medicine.

Key words: nanopore; single molecule level; signal recognition; feature information

EEACC: 7230

doi: 10.3969/j.issn.1004-1699.2023.08.004

纳米孔传感信号的识别和信息处理研究 *

曾祥杰, 王润雨, 罗星月, 武灵芝*, 翁丽星

(南京邮电大学地理与生物信息学院, 江苏 南京 210023)

摘要: 随着纳米孔传感器的发展, 对不同目标分子的信号识别和信息提取的需求越来越大, 如何高效地处理这些强噪音背景下的微弱电信号, 成为纳米孔技术发展的关键问题。因此, 基于 MATLAB 设计了一款自适应的纳米孔脉冲信号的特征事件识别和信息提取的可视化程序, 有助于推动纳米孔大数据信息化处理。通过分析窗口移动对数据分段进行脉冲信号的识别, 同时对选取的数据进行拟合和校正, 减少背景噪音干扰, 提高分辨率。一旦脉冲信号判定为纳米孔分子特征事件, 该软件进一步将特征事件的滞留时间, 信号幅度, 峰位面积等参量进行提取并保存输出。所设计的程序为纳米孔传感用于信息组学的发展和精准医疗提供技术支持。

关键词: 纳米孔; 单分子水平; 信号识别; 特征信息

中图分类号: TP212.3

文献标识码: A

文章编号: 1004-1699(2023)08-1195-06

作为一种高精度的单分子检测技术, 纳米孔越来越广泛地应用于基因测序和生物传感领域, 具有无需标记和扩增、低成本、高灵敏和高通量等优势, 可以识别不同生物分子, 如核酸、蛋白质以及病毒、纳米颗粒等^[1-4]。纳米孔检测方法简单, 一般情况下, 纳米孔传感器是将具有纳米孔结构的薄膜置于电解质溶液中, 当在溶液中施加不同的偏置电压, 溶液中的离子通过纳米孔定向运动, 产生一定的离子电流。当检测分子进入纳米孔, 会引起离子电流的波动。如在适当实验条件下(如孔的大小, 离子强度和所加的电势差), 单个生物分子在孔内取代了足够体积的电解质溶液, 引起可测量的脉冲电流波

动, 能够被高增益放大器记录下来^[5-6]。这些电流变化的强度, 时间, 频率以及波形能够提供生物分子在溶液中体积、浓度、表面电荷以及动态构象变化等信息, 因此对纳米孔中分子易位引起的脉冲信号的识别与信息提取十分关键^[7-10]。

目前纳米孔电信号的处理软件较少, 一般仅适用于与之匹配的仪器, 信息的读取和输出受到限制。特别是随着目前纳米孔传感器检测范围不断扩大, 不同分子易位引起的信号变化更加多样和复杂, 而商业化软件功能单一, 适用范围有限, 对于背景噪音大, 多变信号的判定和处理都存在困难, 多需要手动判断与剔除干扰信号, 不能适应目前纳米孔传感器

项目来源: 国家重点研发计划项目(2017YFA0205300); 江苏省自然科学基金项目(BK20212012); 南京邮电大学项目(NY220222)

收稿日期: 2022-06-15

修改日期: 2022-08-17

的快速发展。为了高效地处理纳米孔大数据,扩大纳米孔的应用范围,很多研究小组都尝试研发纳米孔微弱电信号的信息提取方法。如 Dekker 小组^[11-14]研究了纳米孔离子电流的背景噪音和漂移问题,提出一种局域阈值的迭代运算,使基线电流的平均值及其标准差收敛,更好地确定基线电流。同时 Kim 组^[10,15]通过算术平均、线性拟合、高斯平滑和回归混合等拟合方法处理纳米孔电流的阶梯变化和陡变,进行基线的精准拟合,进一步提取分子过孔事件。这些工作改善了传统方法通过平均整个电流轨迹来确定开孔电流基线的限制,提高了纳米孔信号识别的精度。针对纳米孔信号噪音滤波带来的信息丢失, Pedone 等^[16]提出了一种改进的脉冲信号宽度和高度评估方法,很好地部分补偿了脉冲滤波所造成的误差。同时 Long 组也提出了隐马尔可夫模型和二阶差分的方法来处理纳米孔信号的噪音和滤波问题^[17-18],在一定程度上平衡了信号去噪与失真的问题,但是缺乏普适性。目前机器学习在纳米孔信号处理中的应用也越来越多,如 Zhang 等^[8,19-20]基于深度学习的双路径网络通过训练可以更准确地获取数据信息,甚至在脉冲电流信噪比为 1 时也适用。机器学习具有强大的信号处理前景,不过机器学习算法的适用性受到训练数据集可访问性的限制,同时对用户操作有一定要求。

因此,为了高效准确地处理纳米孔信号,基于自主搭建的纳米孔传感平台,设计了一款适应性强且高通量的纳米孔数据处理可视化软件,可将生物传感器采集的微弱生物电信号数据进行多标准、大批量的信息提取。该软件主要基于 MATLAB 设计了自适应的传感器电信号的分子特征事件识别和信息提取程序,通过设置信号移动窗口范围及双阈值判定标准,对选取的数据段进行脉冲信号的识别,同时对数据进行拟合和校正,排除背景噪音等干扰。一旦纳米孔传感器脉冲信号确定,该软件进一步将特征事件的滞留时间、信号脉冲幅度以及峰位面积等参量进行提取并保存输出,便于特征事件各个参量的统计分析,有利于全面地揭示分子过孔的动力学过程。

1 纳米孔传感器脉冲信号识别程序

如图 1 所示,该程序基于 MATLAB 设计,运行主程序,加载数据,设置参数,进行脉冲信号的信息识别。程序分为分子事件判定,分子事件校正以及信息提取和结果展示三个模块。该程序的设定参数主要有移动窗口宽度和双阈值的设定。首先基线检

测主要是通过局域窗口设定来进行的,根据基线的稳定性来设定移动窗口的宽度。阈值的设定是在检测范围内捕获更多有效的分子过孔事件,主要通过分析窗口内基线电流和信号噪声水平来设定。通过对移动窗口内阈值的初步设定,再借助上升时间和差分函数方法拟合校正,更好地确定一个分子过孔事件的起始点,从而确定脉冲信号事件。一旦脉冲信号被判定为纳米孔分子特征事件,该方法进一步将特征事件的滞留时间、信号幅度,峰位面积等参量进行提取并输出。该方法具有自适应性,多标准对脉冲信号进行判定和校正,保证了信息提取的准确可靠。同时通过对特征信号的滞留时间、幅值以及信号面积等多参量进行提取分析,多维直观地建立纳米孔脉冲信号的信息图谱。该方法精度高,数据为批量化处理,实现了对纳米孔传感器大数据的高通量分析,节约了时间成本,有助于纳米孔传感器的广泛应用。

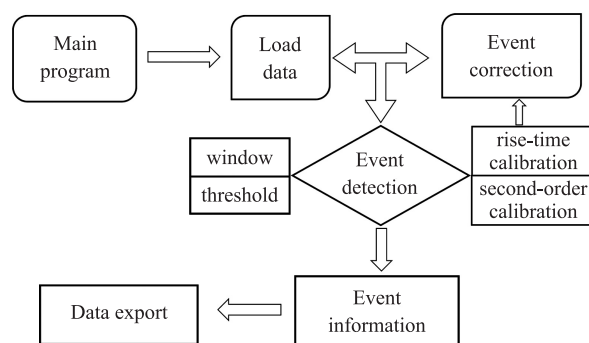


图 1 纳米孔传感器分子脉冲信号识别的程序设计

该程序的界面如图 2 所示。基于 MATLAB 可视化的操作界面,采用窗口移动对局域数据的阈值进行精准设定,对纳米孔检测电流的脉冲信号进行识别和信息提取分析。首先在 Data list 模块对文本数据进行加载,具有很好的兼容性。然后在 Event detection 模块进行参数设置,该过程主要包括阈值

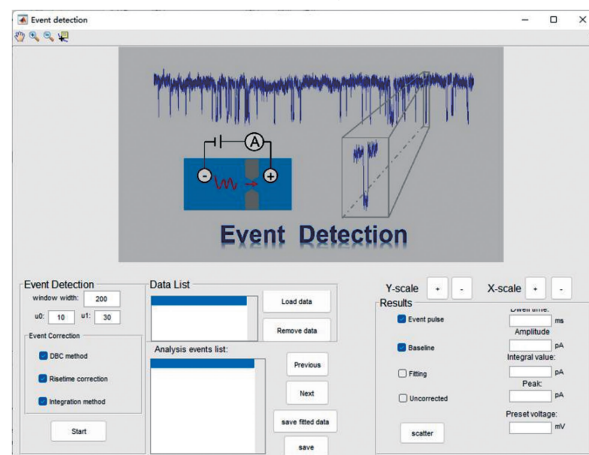


图 2 纳米孔电信号识别主程序界面

设定和数据校正。阈值的设定是在检测范围内捕获更多有效的分子过孔事件,主要通过分析窗口内基线电流和信号噪声水平来设定,排除过大或者过小的异常信号干扰,提高了信号识别的精度。同时对数据进行拟合和校正,排除背景噪音等干扰。接下来在 Result 模块进行信息提取和结果展示等。当纳米孔传感器的电流脉冲信号确定,可以进一步将特征事件的滞留时间、信号脉冲幅度以及峰位面积等参量提取和输出,便于进一步分析。同时该软件还可以进行数据坐标轴的调整,可视化地展示分子事件特征信息和统计分析图谱。

1.1 分子事件判定

当施加一定的偏置电压,纳米孔传感器获得平稳的离子电流,一旦有检测分子,电信号会发生阻塞,形成一个个脉冲信号。每个脉冲信号代表了纳米孔内分子通过时引起的电流变化,称为一个分子易位事件。通过脉冲信号的幅值、时间等信息可以推断出检测分子的种类,物理化学性质等。将纳米孔看成是一个电学性质稳定的元器件,判断分子事件的基本原理就是观察变化的电流信号相对于基线电流的波动幅度。但由于脉冲信号的速度极快,其采样频率为 100 kHz 以上,长时检测中,基线电流可能会有增加或减少的变化,但在较短的一段时间内基线电流是一个十分稳定的值,称为一个局域信号窗口。可以把长时间记录的电流信号段分成一个个窗口,整个信号数据段在分析窗口的移动下进行事件的检测与分析,具有较高的自主适应性,提高了分析效率和准确性。

1.2 分子事件的校正

通过实验采集到的纳米孔电信号一般湮没在强噪音背景中,信噪比较差,所以需要信号数据进行进一步校正。为了更好地使用数学函数进行分析,需要先一步对原始的纳米孔电信号数据进行平滑拟合,以便于后续的校正过程。本程序主要利用 MATLAB 软件中的傅里叶级数拟合函数,对不同长度的信号数据进行傅里叶级数拟合,将上下波动的原始事件信号数据拟合成平滑的信号数据,相当于低通滤波器的作用,使信号数据变得平滑,排除噪声干扰,便于信号分析。拟合后得到的信号数据,通过上升时间和二阶差分等校正函数进行校正分析,进一步消除了背景干扰,提高了信息识别的高效性和准确性。

1.3 脉冲信号特征信息提取

目前通用的纳米孔信号特征通常包括滞留时间与信号幅值(相对于基线电流的幅值,可以看作由

于过孔引起的离子电流变化量),其意义在于分子种类不同,其在纳米孔中的体积排阻效应和动力学过程均不同,通过对大量的分子事件的电信号进行统计分析,就能够总结出每种分子过孔时的电信号特征。为了丰富地刻画分子独特的形貌和表面荷电特性以及在微纳限域空间的运动动力学过程,除了事件信号滞留时间与离子电流变化幅值,该程序设计了多个信号指标进行比对分析,进一步丰富了分子过孔的动力学信息。一旦纳米孔传感器的电流脉冲信号确定,该方法将特征事件的滞留时间、信号脉冲幅度以及整个事件的信号积分面积等参量进行提取并进行保存输出,并提供分子事件特征信息的统计和图谱分析。如离子电流变化峰值对辨识分子种类有一定的对比作用,同时也可以评估外界环境或纳米孔的稳定性;分子事件峰位面积相当于离子电流变化量与事件信号持续时间的乘积,可以对事件进行分级,有助于更精准地判断分子种类,更深层次地分析检测分子形态和结构的变化。

2 金纳米颗粒信号分析

2.1 金纳米颗粒过孔信号检测

随着纳米孔技术的发展,除了 DNA 测序,纳米孔作为一种单分子传感器,广泛用于蛋白质、糖类、脂类以及病毒颗粒等的检测。这些不同大小、形貌结构的检测对象使纳米孔传感信号更加复杂多样。其中金纳米棒就是一个典型的各向异性的纳米结构。通过化学合成,金纳米棒具有可控的长径比,其结构形貌类比纤维状蛋白和杆状的病毒颗粒,为不同形貌生物分子的识别提供了理论参考。通过 MATLAB 设计的纳米孔传感器脉冲信号识别程序,以金纳米棒为例进行纳米孔的信号分析。类比 DNA 的线性结构,金纳米棒是一种各向异性的纳米颗粒,具有可控的尺寸结构,凭借优良的光电特性在生物检测、细胞成像等领域广泛应用。如图 3 所示,以金纳米棒为模型,进行纳米孔检测,对生物医学中

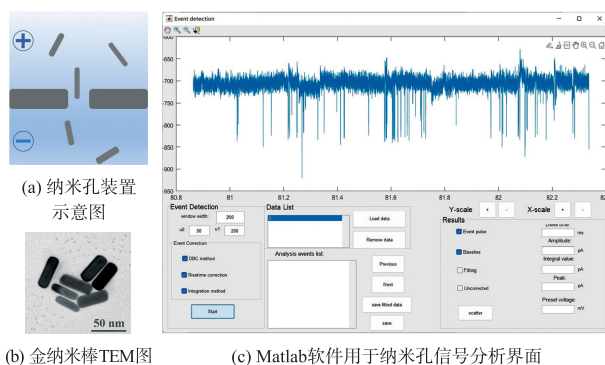


图 3 金纳米棒在纳米孔中信号检测

不同形貌的蛋白质、病毒等生物颗粒的检测都具有重要的借鉴意义。

图 3(a) 为固态纳米孔传感器检测金纳米棒的原理示意图, 纳米孔直径约为 (60 ± 3) nm, 绝缘的氮化硅膜片会把检测的电解质溶液分成 cis 侧和 trans 侧两部分。图 3(b) 为研究中使用的金纳米棒, 通过种子生长法制备, 表面覆盖十六烷基三甲基溴化铵离子层, 带正电, 直径约 22 ± 2 nm, 长度约 50 ± 2 nm, 长径比约为 2.5, 在溶液中具有良好的分散性。当施加偏置电压时, 电解质溶液中的离子在纳米孔道内定向运动形成基线电流。当金纳米棒在电场力驱动下进入到纳米孔道时部分阻断离子通道, 改变了通道内的总体离子浓度, 从而产生了可识别的电流脉冲信号轨迹图。图 3(c) 为 MATLAB 软件用于纳米孔信号分析界面。通过窗口内进行参数设置, 可以实现对金纳米棒易位信号的多维度分析图谱。

2.2 纳米孔信号分析

当溶液中加入金纳米棒后, 记录的离子电流轨迹图中出现了尖刺状的脉冲电流信号。通过程序对不同电压 (400 mV, 600 mV, 900 mV) 下检测到的脉冲电流信号进行统计分析, 对识别的过孔事件进行脉冲幅值 (Amplitude) 和滞留时间 (dwell time) 的直方统计图, 见图 4。

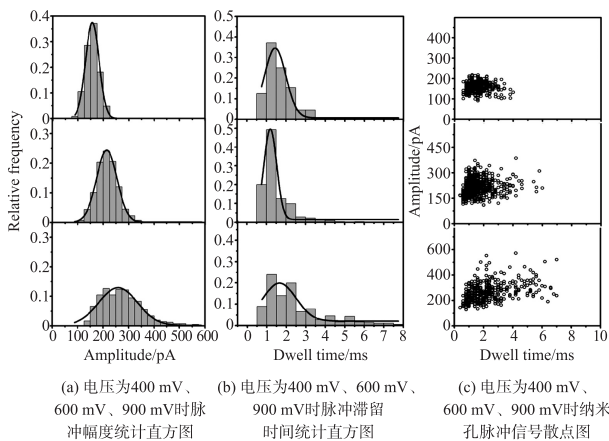


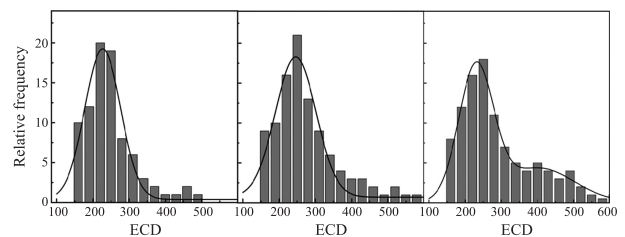
图 4 金纳米棒在纳米孔传感器中不同电压下的电流信号统计图

图 4(a) 从上到下为 400 mV、600 mV、900 mV 对脉冲信号幅值统计分布直方图, 通过高斯拟合可以看出, 电压在 400 mV、600 mV、900 mV 时, 幅值峰值分别 (151 ± 3) pA, (208 ± 4) pA, (289 ± 6) pA, 脉冲幅值随电压的增大而增大, 符合欧姆定律。图 4(b) 为脉冲电流信号的滞留时间统计分布直方图, 其峰值分别为 (1.44 ± 0.047) ms, (1.28 ± 0.027) ms, (1.61 ± 0.150) ms, 有趣的是在 900 mV 下金纳米棒易位信号的滞留时间相对于 600 mV 下信号的滞留时间反

而增加。脉冲信号的滞留时间反映了金纳米棒在纳米孔道内的易位速度。理论上在电压 900 mV 下纳米孔通道内的电场强度要高于 600 mV 时的电场强度, 更快地驱动纳米颗粒过孔, 但实际实验中金纳米棒在孔内的过孔速度减慢, 从而引起滞留时间增加。同时在图 4(c) 中看出, 不同电压下纳米孔事件信号的散点图在不同电压下, 也发生变化。在 900 mV 下, 颗粒过孔事件的散点分布具有更宽的滞留时间和幅值范围, 这表明在高压下随着纳米孔易位事件的增多, 金纳米棒过孔的动力学过程具有复杂性和多样性。

纳米孔传感信号的主要信息包含在脉冲信号的幅值和滞留时间中, 通过幅值的变化, 可以更好地反映过孔颗粒的大小形貌等特征, 而滞留时间也和颗粒的表面电荷等紧密相关^[1,6,8]。像 λ -DNA 线性结构过孔时, 同时伴有折叠和局域折叠等构象, 每个分子事件具有不一样的脉冲电流幅值和滞留时间, 但是这些事件的电流时间积分 (Event Charge Deficit, ECD), 即 $\int \Delta I(t) dt = \text{ECD}$ 都符合一个常数^[1,21], 即不管 λ -DNA 以何种构象通过纳米孔, 只要单个分子长度是一样的, 其阻塞电流事件阻塞部分的面积应该为一个常数。

图 5(a) 中展示了金纳米棒在不同电压下通过纳米孔时的 ECD 统计分布图。在电压为 400 mV、600 mV、900 mV 时, ECD 统计的主峰位都是在 250 附近, 表明大多数纳米棒在电场驱动下, 以相同或相近的形式顺利过孔。但在 900 mV 时, ECD 的直方统计图在主峰位外, 在 ECD = 400 处还有一个小的肩峰, 因此金纳米棒的过孔过程还存在不同相位形



(a) 金纳米棒在不同电压下的电流信号 ECD 统计图
其中从左到右分别是电压为 400 mV, 600 mV, 900 mV

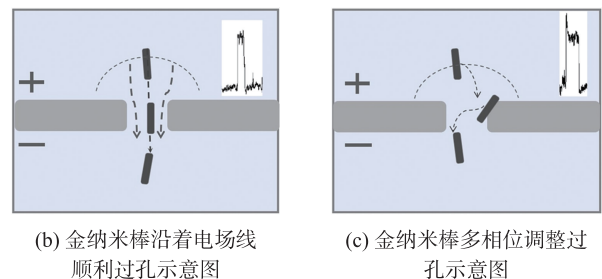


图 5 金纳米棒通过纳米孔

式。通过研究发现,由于金纳米棒表面带正电,同时具有各向异性结构,进入纳米孔口时,多数金纳米棒沿电场线方向,以平行于纳米孔中心轴的相位快速通过纳米孔道,见图 5(b)。但当高压下,金纳米棒在快速进入纳米孔时,会出现以多种相位形式进入纳米孔,与纳米孔表面发生吸附碰撞,同时由于金纳米棒的相位与电场方向不一致,需要相位调整,将会明显比平行相位进入纳米孔道受到的阻力大的多,易位速度减慢,因此在纳米孔道内将会有较长的滞留时间,见图 5(c),这也与金纳米棒在高压下过孔时滞留时间反而延长结果一致。因此通过 MATLAB 程序对信号进行多方位分析,可以更好地揭示金纳米棒各向异性的易位现象,对纳米孔用于各种具有矢量特征的蛋白质、病毒颗粒等的检测提供了理论依据和技术支持,具有十分重要的诊疗意义。

3 总结

纳米孔分子器件凭借独特的电导脉冲原理,在 DNA 测序以及生物分子传感领域具有明显优势,还将为 DNA 数据存储和信息读取等领域带来新的技术革命。不过目前实际应用中,纳米孔传感信号的识别和处理仍然存在限制,如电信号会受到非理想器件噪声以及易位多样性引起的时空分辨率不足等挑战,且很多纳米孔测序数据的分析开发仍比较少,缺少高效的计算方法和分析工具。因此建立准确科学的纳米孔信号的分析方法和技术对纳米孔传感器的发展至关重要。在前期研究的基础上,本研究设计开发了纳米孔传感器电流脉冲信号的识别和信息提取程序,多方位实现了对检测到的纳米颗粒易位事件引起的特征信号的信息识别、统计分析等功能,有利于纳米孔在基因组学、病毒等领域的高灵敏检测,推动精准医学的快速发展。

参考文献:

- [1] Lastra L S, Bandara Y M N D Y, Nguyen M, et al. On the Origins of Conductive Pulse Sensing Inside a Nanopore [J]. *Nature Communications*, 2022, 13(1): 2186.
- [2] Brinkerhoff H, Kang ASW, Liu J, et al. Multiple Rereads of Single Proteins at Single-Amino Acid Resolution Using Nanopores [J]. *Science*, 2021, 374(6574): 1509–1513.
- [3] Arima A, Harlisa IH, Yoshida T, et al. Identifying Single Viruses Using Biorecognition Solid-State Nanopores [J]. *Journal of the American Chemical Society*, 2018, 140(48): 16834–16841.
- [4] Horiguchi Y, Naono N, Sakamoto O, et al. Methodology to Detect Biological Particles Using a Biosensing Surface Integrated in Resistive Pulse Sensing [J]. *ACS Applied Materials & Interfaces*, 2022, 14(17): 20168–20178.
- [5] Zhang S L, Huang G, Versloot R C A, et al. Bottom-up Fabrication of a Proteasome-Nanopore that Unravels and Processes Single Proteins [J]. *Nature Chemistry*, 2021, 13(12): 1192–1199.
- [6] Chuah K, Wu Y, Vivekchand S R C, et al. Nanopore Blockade Sensors for Ultrasensitive Detection of Proteins in Complex Biological Samples [J]. *Nature Communications*, 2019, 10(1): 2109.
- [7] Meyer N, Janot J M, Torrent J, Balme S. Real-Time Fast Amyloid Seeding and Translocation of α -Synuclein with a Nanopipette [J]. *ACS Central Science*, 2022, 8(4): 441–448.
- [8] Chen Y W, Dematties D, Zhang S L. A Guide to Signal Processing Algorithms for Nanopore Sensors [J]. *ACS Sensors*, 2021, 6(10): 3536–3555.
- [9] Shekar S, Chien C C, Hartel A, et al. Wavelet Denoising of High-Bandwidth Nanopore and Ion-Channel Signals [J]. *Nano Letters*, 2019, 19(2): 1090–1097.
- [10] Bandara Y M N D Y, Saharia J, Karawdeniya B, et al. Nanopore Data Analysis: Baseline Construction and Abrupt Change-Based Multilevel Fitting [J]. *Analytical Chemistry*, 2021, 93(34): 11710–11718.
- [11] Smeets R M M, Keyser U F, Dekker N H, et al. Noise in Solid-State Nanopores [J]. *Proceedings of the National Academy Science*, 2008, 105(2): 417–421.
- [12] Fragasso A, Pud S, Dekker C. $1/f$ Noise in Solid-State Nanopores is Governed by Access and Surface Regions [J]. *Nanotechnology*, 2019, 30(39): 395202.
- [13] Alessio F, Sonja S, Cees D. Comparing Current Noise in Biological and Solid-State Nanopores [J]. *ACS Nano*, 2020, 14(2): 1338–1349.
- [14] Plesa C, Dekker C. Data Analysis Methods for Solid-State Nanopores [J]. *Nanotechnology*, 2015, 26(8): 084003.
- [15] Saharia J, Bandara Y M N D Y, Goyal G, et al. Molecular-Level Profiling of Human Serum Transferrin Protein Through Assessment of Nanopore-Based Electrical and Chemical Responsiveness [J]. *ACS Nano*, 2019, 13(4): 4246–4254.
- [16] Pedone D, Firmkes M, Rant U. Data Analysis of Translocation Events in Nanopore Experiments [J]. *Analytical Chemistry*, 2009, 81(23): 9689–9694.
- [17] Zhang J, Liu X, Ying Y L, et al. High-Bandwidth Nanopore Data Analysis by Using a Modified Hidden Markov Model [J]. *Nanoscale*, 2017, 9(10): 3458–3465.
- [18] Gu Z, Ying Y L, Cao C, et al. Accurate Data Process for Nanopore Analysis [J]. *Analytical Chemistry*, 2015, 87(2): 907–913.

- [19] Dematties D, Wen C, Pérez M D, et al. Deep Learning of Nanopore Sensing Signals Using a bi-Path Network [J]. ACS Nano, 2021, 15(9): 14419–14429.
- [20] Tsutsui M, Takaai T, Yokota K, et al. Deep Learning-Enhanced Nanopore Sensing of Single-Nanoparticle Translocation Dynamics [J]. Small Methods, 2021, 5(7): e2100191.
- [21] Fologea D, Gershow M, Ledden B, et al. Detecting Single Stranded DNA with a Solid State Nanopore [J]. Nano Letters, 2005, 5(10): 1905–1909.



曾祥杰(1997—),男,南京邮电大学硕士研究生,主要研究方向为纳米生物传感,1020173014@njupt.edu.cn;



武灵芝(1981—),女,副教授,南京邮电大学硕士生导师,主要研究方向为生物信息与生物传感,wulz@njupt.edu.cn。