

封面文章

基于无监督学习方法的细胞膜内单分子扩散运动分析:
胆固醇对模型膜和活细胞膜流动性的不同影响*谭金鹏^{1)2)#} 张婉婷^{2)3)#} 徐成^{2)#} 卢雪梅⁴⁾⁵⁾朱文圣¹⁾ 杨恺^{3)5)†} 元冰^{2)‡}

1) (东北师范大学数学与统计学院, 长春 130024)

2) (松山湖材料实验室, 东莞 523808)

3) (苏州大学物理科学与技术学院, 软凝聚态物理及交叉研究中心, 苏州 215006)

4) (苏州城市学院光学与电子信息学院, 苏州 215104)

5) (苏州大学, 前沿材料物理与器件省高校重点实验室, 苏州 215006)

(2024 年 7 月 2 日收到; 2024 年 7 月 31 日收到修改稿)

单分子运动追踪是研究软物质体系尤其是生命体系动力学过程和分子相互作用的重要方法, 但如何理解生命体系中单分子运动行为的复杂性仍是一个巨大的挑战. 针对这一问题, 本工作提出了一种可对单分子轨迹进行高效识别和分类的、基于无监督学习的“两步归类法”: 首先利用熵约束最小二乘法对扩散轨迹的受限程度进行区分, 继而通过统计检验将非受限轨迹划分为亚扩散、正常扩散和超扩散等不同运动模式类型. 利用该方法, 本工作解析了 DOPC 模型细胞膜和活细胞膜内的单分子扩散运动特征, 揭示了胆固醇成分对二者的差异影响. 结果显示: 模型膜和活细胞膜均包含多种不同的扩散模式; 在 DOPC 模型膜体系中, 胆固醇成分会阻碍膜内的分子扩散运动, 且阻碍程度与胆固醇含量正相关; 在活细胞体系中, 分子运动速率显著低于模型膜体系, 并且, 胆固醇的去除会进一步减慢分子扩散速率. 本研究有助于从单分子运动角度深入理解生物分子运动行为的复杂性及其对体系环境的依赖性.

关键词: 单分子运动, 无监督学习, 模型细胞膜, 胆固醇**PACS:** 87.16.D-, 87.15.K-, 87.15.H-, 87.15.Vv**DOI:** 10.7498/aps.73.20240915

1 引言

细胞膜的流动性是其重要特征之一, 与许多基本生理过程密切相关, 包括细胞迁移、胞吞、胞吐、信号识别与传导以及细胞对外界刺激的响应等^[1].

膜的流动性受到细胞膜组分多样性和膜内环境拥挤性等因素的强烈影响, 并呈现出与功能相关的空间异质性 (从分子水平上形成的脂筏到中尺度水平上形成的微区) 以及时间动态特征^[2,3]. 这些复杂多变的特点给常规生化研究方法带来困难^[4]. 因此, 如何从分子层次上理解膜流动性仍然是目前相关

* 国家自然科学基金 (批准号: 12274307, 32230063, 12347102, 22307090, 22203059)、广东省基础与应用基础研究基金 (批准号: 2023A1515011610)、江苏省自然科学基金 (批准号: BK20210100) 和前沿材料物理与器件省高校重点实验室 (苏州大学) (批准号: KJS2131) 资助的课题.

同等贡献作者.

† 通信作者. E-mail: yangkai@suda.edu.cn

‡ 通信作者. E-mail: yuanbing@sslab.org.cn

研究面临的主要挑战^[5-7].

胆固醇 (cholesterol, Chol) 是调节细胞膜结构、流动性、渗透性、张力以及对外界刺激响应性的关键分子, 占哺乳细胞膜内脂质总量的约 40%^[8]. 从结构上看, 胆固醇会嵌入脂质膜并与磷脂共同组装, 并限制其酰基链构象灵活性^[9]. 此外, 胆固醇与不同酰基链长度和饱和度的脂质之间相互作用的差别会导致形成具有特定组成、堆积密度和有序性的相态 (如液体有序相 Lo 和液体无序相 Ld 等). 然而, 目前人们对于胆固醇对膜流动性的影响仍知之甚少^[10,11], 部分研究结果存在广泛争议^[7]. 由于活细胞膜的复杂性和现有研究方法的局限性, 关于胆固醇对膜性质影响的研究常使用由不饱和脂质 1, 2-二油基-*n*-甘油-3-磷脂酰胆碱 (DOPC) 或饱和脂质双棕榈酰基磷脂酰胆碱 (DPPC) 和胆固醇所组成的模型膜来进行^[5]; 但值得注意的是, 基于模型细胞膜与真实活细胞膜的研究结果之间可能存在差异^[12,13]. 一种可能性是, 尽管胆固醇局部增强了不饱和及饱和脂质膜的曲率刚度, 但其却能促进液体有序 (Lo) 区域中的流动性和/或阻碍液体无序 (Ld) 相区域中物质扩散^[10,14].

单分子运动追踪技术 (SMT) 为监测细胞内尤其是细胞质膜界面处的动力学过程和分子相互作用提供了一种有效方法^[15]. SMT 已被广泛应用于多个生物过程的研究中, 包括外源颗粒 (如病毒和 mRNA) 的内化、膜蛋白的激活以及跨膜信号传导 (如配体诱导的受体内吞作用) 等^[16,17]. 目前, 单分子技术主要侧重于颗粒和膜蛋白的运动跟踪, 而对于细胞膜的关键组成部分——脂质的扩散行为则关注较少^[18]. 实际上, 单分子运动追踪技术既专注于研究单个分子层面的信息, 又通过大量分子提供系综层面分析, 这为多尺度解析脂质分子运动特征及其变化提供了可能的有效途径^[19,20].

生物体系中的单分子运动行为极为复杂, 例如, 粒子在活细胞环境中会表现出正常扩散、异常扩散以及受限运动等多种运动模式. 因此, 如何准确识别粒子在复杂环境中的运动模式是当前单分子运动研究的一个重点. 近年来, 不同的机器学习方法被尝试用于单分子运动轨迹的识别, 如逻辑回归^[21]、随机森林^[22]、贝叶斯方法^[23]、深度学习^[24]等. 相较传统方法, 这些方法展示出更高的识别准确性, 但也存在一些局限. 例如, 以往的一些方法集中于监督学习方法, 通过对数据进行处理将单分

子轨迹识别问题转化为分类问题. 然而真实数据的标注往往较为困难, 因此这类方法常基于计算模拟生成的训练数据集来进行模型训练, 然而模拟数据与实验数据的区别会导致模型可能存在着较大的局限性^[25].

本研究采用单脂质分子追踪技术, 探究了胆固醇对模型膜和活细胞膜流动性的差异影响. 特别是, 我们开发了一种基于无监督学习方法的分类机器学习模型, 能有效地对细胞膜内单分子扩散运动进行分类, 并针对不同类型的运动模式分别进行分析; 在此基础上, 研究了由 DOPC 和胆固醇组成的模型膜以及活细胞膜中的脂质分子运动特征, 发现去除胆固醇对这两种体系会产生不同的影响. 本研究为生物体系中单分子运动追踪技术提供了一种定量且有效的轨迹分类分析方法, 并从单分子运动角度阐明了胆固醇对不同成分膜流动性所带来的差异影响.

2 生命体系中单分子扩散运动行为的特征

分子扩散运动是分子间相互作用的一种表现形式; 细胞膜内脂质分子扩散行为的变化可反映出细胞膜成分、结构和性质的变化, 甚至是界面生物过程的演变. 人们在实验中广泛观察到分子在生物相关环境中的扩散运动与其在稀溶液中的情况发生偏离. 这通常归因于细胞和细胞膜的环境拥挤性和结构非均质性等因素的影响, 最明显体现为均方位移 (MSD) 随滞后时间 (lag time) 表现出次线性幂律增长规律、扩散系数 (D_L) 显著降低、空间位移 (displacement) 非高斯分布性、以及存在非均匀扩散模式和静止粒子等^[19,26-28]. 目前人们已经提出了几种理论模型来解释这些现象, 如高斯模型、分数布朗运动模型 (fBM)、连续时间随机行走模型 (CTRW)、以及描述异质环境中阻碍运动的洛伦兹模型等^[29]. 然而, 在不同体系中关于脂质分子扩散行为特征、胆固醇对其影响以及模型膜与真实细胞体系之间的差异性仍不清楚.

目前, 粒子运动类型推断主要是基于均方位移 (MSD) 分析^[30]:

$$\text{MSD}(t) = E \left(\|X_{t+t_0} - X_{t_0}\|^2 \right), \quad (1)$$

其中, X_t 表示粒子运动轨迹, t_0 表示轨迹起始时

刻, t 表示轨迹运动时间, $\|\cdot\|$ 表示欧几里得范数, E 表示随机变量的期望. 由于真实实验中分子运动轨迹数量有限, 通常会使用由单个轨迹计算出的时间平均均方位移 (TAMSD) 作为 MSD 的估计:

$$\langle r^2(n) \rangle = \frac{1}{N-n} \sum_{i=1}^{N-n} \|X_{t_{i+n}} - X_{t_i}\|^2, \quad n = 1, 2, \dots, N-1, \quad (2)$$

其中 n 表示滞后项. 值得强调的是, 滞后项的选择会影响 MSD 的计算. 根据以前的研究结果^[31,32], 我们将以不超过粒子轨迹长度一半 ($N/2$) 作为滞后项的选择标准^[33].

根据 Saxton 和 Jacobson^[34] 的研究结果, 不同扩散类型具有不同的 MSD 曲线形状 (图 1(a)), 其函数形式如下:

$$\hat{r}^2(t) = 4Dt, \quad \text{正常扩散}, \quad (3)$$

$$\hat{r}^2(t) = 4Dt^\alpha, \quad \text{异常扩散}, \quad (4)$$

$$\hat{r}^2(t) = 4Dt + (vt)^2, \quad \text{定向扩散}, \quad (5)$$

$$\hat{r}^2(t) \simeq r_c^2 \left[1 - A_1 \exp\left(-\frac{4A_2Dt}{r_c^2}\right) \right], \quad \text{受限扩散}. \quad (6)$$

这里 D 表示粒子的扩散系数. α 是表示异常扩散指数: 当 $0 < \alpha < 1$ 时, 粒子表现为亚扩散 (sub-diffusion); 当 $1 < \alpha < 2$ 时, 粒子表现为超扩散 (super-diffusion); 当 $\alpha = 1$ 时, 则为正常扩散 (normal

diffusion). v 表示粒子定向扩散时的速度. 方程 (6) 中 r_c 表示约束半径, 常数 A_1 和 A_2 为约束形状参数. 如图 1(a) 所示, 不同扩散类型的 MSD 函数图像会呈现出独特的特征, 这为采用不同的函数族来识别扩散运动类型提供了依据. 本文主要讨论粒子在二维空间的运动. Kusumi 等^[35] 和 Saxton^[36] 的研究表明二维空间受限一般包括方形和圆形两种情况, 并给出了方程 (6) 在这两种情况下的精确级数解. 在后续的研究中, 我们使用圆形对二维受限区域进行刻画. 并且, 由于我们实验体系中不包含粒子的定向扩散, 因此在下面的分析中主要考虑正常扩散、异常扩散和受限扩散这 3 种运动模式.

3 基于无监督学习方法对细胞膜内单脂质分子扩散轨迹进行分类

我们发展了一种“两步归类法”对单分子轨迹进行识别和分类 (图 1(b)). 与以往的方法不同的是, 这个方法是一种无监督学习方法, 能够在没有额外训练数据的情况下自动对大量单分子轨迹数据进行识别, 具有简单高效等特点. 此外, 该方法考虑了受限运动、正常扩散和异常扩散等多种粒子运动模式, 因而可对复杂粒子运动轨迹进行有效识别和分类, 并计算相应的扩散系数和异常指数. 简而言之, 我们首先建立了一个概率混合模型, 并通过熵约束最小二乘法来识别受限扩散和非受限扩

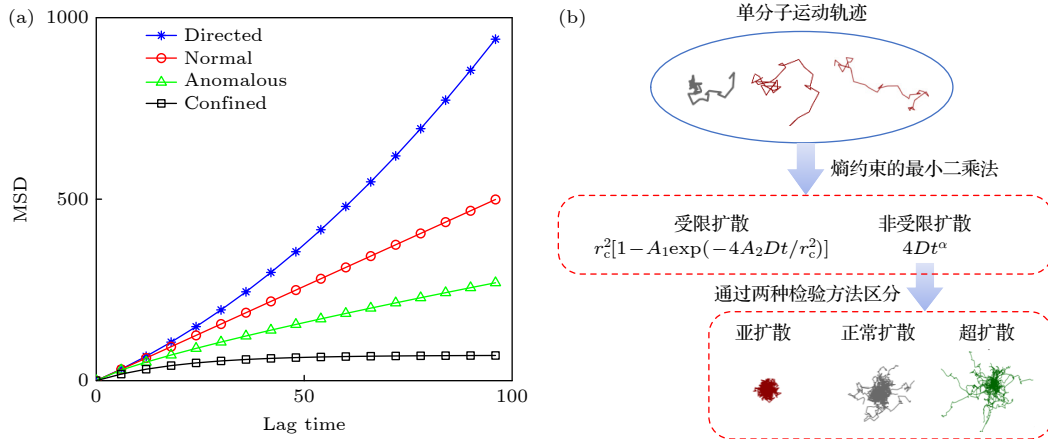


图 1 单分子扩散运动的不同模式 (a) 模拟运动模型的 MSD 曲线, 即定向扩散 (蓝色)、正常扩散 (红色)、异常扩散 (绿色) 和受限扩散 (黑色), 横坐标为滞后时间 (lag time); (b) 采用“两步归类法”对单分子轨迹进行识别和分类的步骤, 第一步区分受限扩散与非受限扩散, 第二步将非受限扩散进一步区分

Fig. 1. Different modes of single-molecule diffusion. (a) Simulated mean square displacement (MSD) curves depicting directed diffusion (blue), normal diffusion (red), anomalous diffusion (green), and confined diffusion (black). (b) The “two-step classification method” employed to identify and classify single-molecule trajectories: the first step involves distinguishing between confined and unconfined diffusion, followed by further differentiation of unconfined diffusion in the second step.

散轨迹; 然后通过统计检验方法确定非受限轨迹的扩散类型 (亚扩散、正常扩散、超扩散).

假设粒子的运动轨迹为受限扩散的概率为 P_1 , 非受限扩散概率为 P_2 , 且 $P_1 + P_2 = 1$, 则粒子运动 TAMSD 可表示为

$$\hat{r}^2(t) = P_1 \cdot r_c^2 \left[1 - A_1 \exp \left(-\frac{4A_2 D t}{r_c^2} \right) \right] + P_2 \cdot 4Dt^\alpha. \quad (7)$$

因此, 当 $P_1 > P_2$ 时, 轨迹以更大的概率属于受限扩散; 当 $P_1 < P_2$ 时, 轨迹以更大的概率属于非受限扩散. 我们可通过熵约束最小二乘方法对参数进行估计:

$$\begin{aligned} \operatorname{argmin}_{P, D, \alpha} & \left\{ \frac{1}{T} \sum_{t=1}^T (\langle r^2(t) \rangle - \hat{r}^2(t))^2 + \lambda H(P) \right\}, \\ \text{s.t.} & \sum P_i = 1, P \geq 0, D > 0, 0 < \alpha < 2, \end{aligned} \quad (8)$$

其中 λ 为惩罚参数, $\lambda \geq 0$, 通过调整 $H(P)$ 的大小控制模型复杂度. P_i 表示概率, 求和为 1. 在理想情况下, P_i 的取值只有一个为 1, 其余为 0, 表示此时的轨迹可被准确识别. 因此, 我们添加信息熵 ($H(P)$) 对模型进行约束, 具体形式为

$$H(P) = \sum_i -P_i \log P_i. \quad (9)$$

信息熵最早由 Shannon^[37] 提出, 用来度量信息不确定性. 这里通过信息熵对概率 P 进行惩罚, 以降低模型复杂度.

由于方程 (7) 是非线性的, 因此 (8) 式的求解是一个非凸非线性优化问题. 序列二次规划算法 (SQP) 是求解约束非线性优化问题最有效的方法之一^[38], 具有收敛性好、计算效率高、边界搜索能力强等优点. 因此, 我们使用 SQP 将原始非线性优化问题转换为多个简单二次规划子问题进行求解. 对于非凸问题求解, 可以使用启发式算法结合 SQP, 如 GA-SQP^[39], PSO-SQP^[40]. 但启发式算法的时间复杂度较高, 不利于处理大量粒子运动轨迹数据. 并且, 非凸问题求解容易受到初值选择的影响, 因此我们通过随机抽样方法产生多组不同的初值, 使用不同初值迭代求解得到多个局部最优解, 并从中选择使 (8) 式达到最小的解作为最终结果. 通过真实数据测试后发现 200 组初值即可在计算精度和时间之间达到较好平衡, 因此, 在后续处理中均设置初值组数为 200.

通过 (8) 式的求解可以区分出扩散轨迹的受限或非受限特征. 之后, 还需要对非受限轨迹中的正常扩散、亚扩散和超扩散等类型进行识别. 本文中识别是基于两种不同统计检验方法来进行, 它们的区别主要在于选择检验统计量不同.

在不同类型扩散中, 异常指数 α 的取值范围不同, 因此可通过异常指数 α 直接对轨迹进行区分, 即设置一个任意截断参数 c ($0 < c < 1$):

- 1) 若 $\alpha < 1 - c$, 则为亚扩散.
- 2) 若 $\alpha > 1 + c$, 则为超扩散.
- 3) 若 $1 - c < \alpha < 1 + c$, 则为正常扩散.

截断参数 c 一般根据经验来判断, 常见取值范围为 0.1—0.25, 但不同取值对分类结果影响较大^[41,42]. 因此给定一个合理 c 值是保证方法有效的关键. 对于异常指数 α 的估计, 一种最常用方法是对 TAMSD 进行对数变换, 建立线性模型,

$$\log(r^2(t_i)) = \log(D) + \alpha \log(t_i). \quad (10)$$

然后通过线性模型最小二乘估计得到关于 α 的估计值. 现有研究已证明在长轨迹下此方法是 α 的渐近无偏估计, 随着轨迹的增加方差趋于 0; 即使对于单个 (足够长的) 轨迹, 这种方法也可准确估计异常指数^[33]. 因此, 我们考虑基于线性模型检验的统计量给出 c 取值:

$$c = t_{\delta/2}(n-2) \operatorname{SE}(\hat{\alpha}), \quad (11)$$

其中,

$$\operatorname{SE}(\hat{\alpha}) = \sqrt{\frac{\frac{1}{n} - 2 \sum (\log(r^2(t_i)) - \widehat{\log}(r^2(t_i)))^2}{\sum (\log(t_i) - \overline{\log(t_i)})^2}},$$

$\widehat{\log}(r^2(t_i))$ 为 TAMSD 的预测值, 具体见文献^[33]; $t_{\delta/2}(n-2)$ 表示自由度为 $n-2$ 的 t 分布, $\delta/2$ 设为 5%. 当异常指数 α 的 90% 置信区间包含 1 时, 则可以认为是正常扩散; 当置信区间上界小于 1 时, 则为亚扩散; 当置信区间的下界大于 1 时, 则为超扩散.

第二种识别方法是基于 Briane 等^[31] 提出的检验统计量 T_N :

$$T_N = \frac{D_N}{\sqrt{(t_N - t_0) \hat{\sigma}_N^2}}, \quad (12)$$

其中, t_0 表示起点时刻, t_N 表示终点时刻, $D_N = \max_{i=1, \dots, N} \|X_{t_i} - X_{t_0}\|_2$ 表示在给定时间段 $[t_0, t_N]$ 内粒子从起点移动的最大距离, T_N 可看作是关于 D_N

的标准化. $\hat{\sigma}_N^2$ 表示 σ^2 的一致估计量, 如:

$$\hat{\sigma}_N^2 = \frac{1}{2N\Delta} \sum_{j=1}^N \|X_{t_j} - X_{t_{j-1}}\|_2^2 \quad (13)$$

或

$$\hat{\sigma}_N^2 = \frac{1}{2N\Delta} \sum_{j=1}^{N-1} \|(X_{t_{j+1}} - X_{t_j}) - (X_{t_j} - X_{t_{j-1}})\|_2^2, \quad (14)$$

其中 $\Delta = t_i - t_{i-1}$, 表示单位位移时间. 对于自由扩散 (布朗运动), T_N 分布不依赖于 σ . 当粒子运动受到阻碍, 运动过程停留在其初始位置附近, 表现为次扩散状态, 此时 T_N 值较小; 当粒子远离其初始位置, 表现为超扩散状态. 例如粒子可以由分子马达驱动沿某个方向移动 (超扩散), 此时对应的 T_N 值较大. 因此, 可借助 T_N 判断粒子的扩散状态. 在文献 [31] 中, Briane 等分析了大样本下 T_N 的渐近分布, 然而真实实验中轨迹长度往往有限, 例如我们在实验中观察到的轨迹长度一般不大于 100, 此时使用渐近分布作为短轨迹样本的分位数近似可能是不够准确的 [41]. 但是, T_N 及其分位数 $q_N(\delta/2)$ 的分布可以通过 Monte Carlo 模拟来近似 [31,41]. 当 $T_N < q_N(\delta/2)$ 时, 运动轨迹为亚扩散, 当 $T_N > q_N(1 - \delta/2)$ 时, 运动轨迹为超扩散, 其他则为正常扩散. 具体计算过程见补充材料 (online) 中算法 S1 和算法 S2.

4 模型膜和活细胞内单分子扩散行为分析

4.1 实验方法

模型膜制备 制备采用以下分子: 1, 2-二油基-sn-丙三基-3-磷脂酰胆碱 (DOPC, Avanti 公司)、头部基团以丽丝胺罗丹明标记的 1, 2-二油基-sn-丙三基-3-磷脂酰乙醇胺 (RhB-DOPE, Avanti 公司)、胆固醇 (Chol, Merk 公司)、甲基- β -环糊精 (Me β CD, Sigma 公司).

以传统的囊泡融合法制备平板支撑脂质双层膜, 其脂质成分分别为纯 DOPC, DOPC 与 Chol 摩尔比为 4:1, DOPC 与 Chol 摩尔比为 1:1, 所有膜内以 0.0001% 比例 (摩尔比) 掺以 RhB-DOPE; 之后, 将 3 种模型膜经 5 mmol/L 浓度 Me β CD 溶液 (溶解在 PBS 中) 在 37 °C 孵育 30 min 后以 PBS 冲洗 (在我们之前的工作中已经证明该方法能够有

效去除模型膜和活细胞内的胆固醇却不会导致膜的破坏或细胞毒性) [43]. 由此获得 6 种模型膜体系.

活细胞样品制备及单分子追踪 小鼠黑色素瘤细胞 (B16) 来自中国科学院典型培养物保藏委员会细胞库 (CACC). 细胞培养于 DMEM 培养基 (Gibco, Thermo Fisher) 中, 添加胎牛血清 (FBS, 10% (体积分数), Gibco) 和青霉素/链霉素 (1% (体积分数), Gibco). 将细胞经 5 mmol/L 浓度 Me β CD 溶液 (溶解于无血清 DMEM 培养基) 在 37 °C 孵育 30 min 后以 PBS 冲洗来去除胆固醇, 之后培养于用无血清培养基中. 在单分子实验前, 将细胞以 CellMask™ 荧光分子 (Thermo; 以 5×10^{-5} mg/mL 浓度溶解于 PBS) 孵育 30 min 后以 PBS 冲洗, 之后立刻置于 TIRFM 下观察. 在基于模型膜的参照实验中, CellMask™ 荧光分子与 RhB-DOPE 荧光脂质表现出相似的单分子扩散运动行为 (补充材料图 S1 (online)).

单脂质分子追踪 将模型膜及活细胞样品置于全内反射荧光显微镜 (TIRFM, Olympus IX71 配以 100 $\times$ 全内反射专用油镜) 下观察, 以 EMCCD (Andor DU-897U) 采集单分子荧光信号. RhB 荧光团的激发使用 532 nm 激光器 (Coherent, 200 mW, 60% 功率), 采集使用 EM BP 585/40 nm. Cell Mask™ 的激发使用 647 nm 激光器, 采集使用 EM BP 706/95 nm. 视频采集频率为每秒 50 帧, 成像尺寸为 41 μ m $\times$ 82 μ m, 分辨率为 160 nm, 每个视频包含 300 帧. 截取 60 帧 (模型膜) 或 20 帧 (活细胞膜) 长度的轨迹进行分析.

4.2 模型膜内的单脂质分子扩散运动分析及胆固醇的影响

采集 3 种成分模型膜 (纯 DOPC, DOPC:Chol = 4:1, DOPC:Chol = 1:1) 及其经 Me β CD 去除胆固醇后共 6 个体系模型膜内的单脂质扩散运动轨迹. 在所有体系中, 脂质轨迹皆显示出显著的异质性, 这可体现在轨迹回转半径 (R_g) 的巨大差异上 (补充材料图 S2 (online)). 基于前文阐述的“两步归类法”对各体系内单脂质分子扩散轨迹进行识别和分类, 将其分为受限扩散 (confined diffusion)、亚扩散 (sub-diffusion)、正常扩散 (normal diffusion)、超扩散 (super-diffusion) 4 种类型, 并由此将轨迹分为 4 个子群 (表 1, 典型曲线显示于补充材料图 S3 (online)). 分别计算不同体系不同子群内

表 1 不同模型膜体系内单脂质分子轨迹分类及相应的系综扩散系数和异常指数数值, 其中轨迹长度为 60 帧
Table 1. Classification of single lipid trajectories in different model membrane systems and corresponding ensemble diffusion coefficient and anomaly index values. The trajectory length is 60 frames.

	Subgroup	Control			+Me β CD		
		DOPC	DOPC:CHOL = 4:1	DOPC:CHOL = 1:1	DOPC	DOPC:CHOL = 4:1	DOPC:CHOL = 1:1
Trajectory number	Confined	282	383	497	293	334	330
	Sub-diffusion	943	1143	1036	947	799	811
	Normal diffusion	318	294	290	305	200	207
	Super-diffusion	667	577	524	647	448	449
Proportion/%	Confined	12.76	15.98	21.18	13.37	18.75	18.36
	Sub-diffusion	42.67	47.68	44.14	43.20	44.86	45.13
	Normal diffusion	14.39	12.27	12.36	13.91	11.23	11.52
	Super-diffusion	30.18	24.07	22.33	29.52	25.15	24.99
Diffusion coefficient $D_L/(\mu\text{m}^2\cdot\text{s}^{-1})$	Confined	0.045	0.032	0.029	0.064	0.034	0.061
	Sub-diffusion	1.086	0.582	0.308	0.805	0.784	0.717
	Normal diffusion	2.421	1.581	0.870	2.100	2.051	1.781
	Super-diffusion	4.149	3.211	1.444	3.908	3.650	3.381
Anomaly index (α)	Confined	0.116	0.094	0.068	0.207	0.099	0.171
	Sub-diffusion	0.709	0.629	0.512	0.629	0.635	0.614
	Normal diffusion	0.909	0.869	0.916	0.948	0.910	0.874
	Super-diffusion	1.133	1.103	1.178	1.141	1.132	1.132

单分子轨迹的系综平均 MSD, 并进一步计算系综扩散系数 (D_L) 和异常指数 (α)(表 1); 同时, 计算单分子轨迹的时间平均 MSD 以及对应的单分子扩散系数 (补充材料图 S3 (online)), 并展示其分布 (图 2).

由分类及计算结果可知, 6 个模型膜体系内皆同时包含 4 种扩散运动类型, 且亚扩散类型轨迹所占比例最高, 约 $45\%\pm 2\%$. 随膜内胆固醇含量增大 (由 0 增至 20%, 50%), 受限运动和亚扩散分子比例增多 (由 55% 增至 65%; 尤其是受限运动分子, 由 13% 增至 21%); 同时, 正常扩散和超扩散运动分子比例降低 (由 45% 降至 35%). 更显著的是, 胆固醇增加会导致各子群的系综扩散系数显著降低 (除受限扩散持续保持低值), 尽管其异常指数并没有明显差异. 这一规律同时体现在系综扩散系数 (表 1) 及单分子扩散系数概率分布 (图 2) 中. 进一步地, 后者表明该变化是由于快速运动分子的速率减慢 (如图 2(c) 中峰 A 的下移) 和慢速运动分子的数目增多 (如图 2(c) 中峰 B 的升高) 这两个因素同时导致的 (典型曲线见补充材料图 S3 (online)). 总之, 上述结果表明胆固醇阻碍了 DOPC 模型膜

内脂质分子扩散, 使得所有分子运动速率变慢, 更造成 10% 的分子由正常扩散或超扩散模式转为亚扩散甚至受限模式.

有趣的是, 经 Me β CD 去除胆固醇之后的 3 个模型膜体系, 它们在 4 个子群中皆表现出相似的分子数目比例、系综扩散系数和异常指数, 并与纯 DOPC 膜情况接近, 例如其正常扩散子群的扩散系数约为 $(2.0\pm 0.1) \mu\text{m}^2/\text{s}$; 这一点在单分子扩散系数分布 (图 2 绿色分布) 和系综 MSD 曲线 (补充材料图 S4 (online) 蓝色箭头) 中也得到了验证. 这意味着 Me β CD 去除胆固醇能够将 DOPC/Chol 混合膜内的脂质分子扩散运动恢复至纯 DOPC 膜的水平.

值得注意的是, 活细胞膜内的单分子扩散运动呈现出与模型膜不同的特征. 由于活细胞膜内复杂拥挤的环境降低了脂质分子发生超扩散运动的可能性, 所以我们将分子轨迹分为受限运动、亚扩散和正常扩散 3 个子群 (表 2 和图 3). 结果显示, 所有分子运动普遍较慢, 其中亚扩散分子比例最高, 约占 45%; 受限运动分子比例约为 31%, 远高于模型膜内 13%—22% 的水平; 正常扩散分子仅占 24%. 同时, 亚扩散和正常扩散分子的系综扩散系

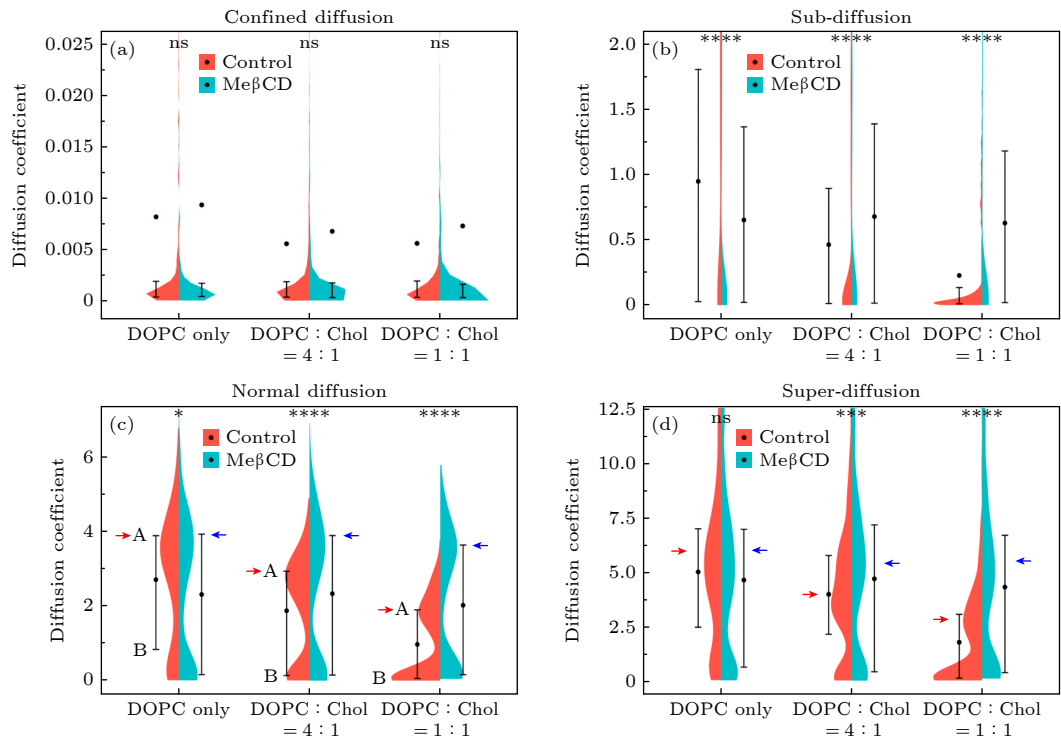


图 2 不同模型膜体系内单分子扩散运动分析, 即利用“两步归类法”将其分为受限扩散 (Confined diffusion)、亚扩散 (Sub-diffusion)、正常扩散 (Normal diffusion)、超扩散 (Super-diffusion) 4 个子群, 分别计算其单分子轨迹的时间平均 MSD 并作出其分布. 轨迹数目见表 1

Fig. 2. Analysis of single lipid diffusion in various model membrane systems using the “two-step classification method”. Trajectories were categorized into four subgroups, namely confined diffusion, sub-diffusion, normal diffusion, and super-diffusion. The time-averaged MSD of individual trajectories was calculated to obtain the distribution. The number of trajectories is presented in Table 1.

表 2 经 MeβCD 去除胆固醇前后活细胞膜内单分子轨迹分类及相应的系综扩散系数和异常指数数值, 其中轨迹长度为 20 帧

Table 2. Classification of single-molecule trajectory in living cell membranes before and after cholesterol depletion by MeβCD, along with corresponding ensemble-averaged diffusion coefficient and anomaly index values. The trajectory length is 20 frames.

	Subgroup	Control	+MeβCD
Trajectory number	Confined	452	304
	Sub-diffusion	645	377
	Normal diffusion	339	103
Proportion/%	Confined	31.48	38.78
	Sub-diffusion	44.92	48.09
	Normal diffusion	23.61	13.14
Diffusion coefficient $D_L/(\mu\text{m}^2\cdot\text{s}^{-1})$	Confined	0.017	0.011
	Sub-diffusion	0.069	0.040
	Normal diffusion	0.465	0.260
Anomaly index (α)	Confined	0.146	0.095
	Sub-diffusion	0.437	0.461
	Normal diffusion	0.983	0.861

数分别为 $0.07 \mu\text{m}^2/\text{s}$ 和 $0.47 \mu\text{m}^2/\text{s}$, 均远低于模型膜对应的数值. 更有趣的是, MeβCD 去除胆固醇进一步降低了活细胞膜内的分子扩散 (补充材料图 S5 (online)). 这一方面体现在受限分子比例增大 (由 31% 到 39%) 和正常扩散分子比例降低 (由 24% 至 13%); 另一方面体现在 3 个子集内系综扩散系数普遍减小 (例如正常扩散分子的扩散系数由 $0.47 \mu\text{m}^2/\text{s}$ 降至 $0.26 \mu\text{m}^2/\text{s}$); 单分子扩散系数概率分布图 (图 3) 及位移概率分布图 (补充材料图 S6 (online)) 显示, 这一变化是由于大量分子运动速率普遍降低 (如图 3(c) 中橙色箭头) 和慢速运动分子比例增大 (如图 3(c) 中峰 A、峰 B 的升高; 图 S6 中蓝色箭头) 共同导致的. 总之, 活细胞膜内单分子扩散运动比 DOPC/Chol 模型膜内慢得多, 存在着更多的受限分子, 且去除胆固醇会进一步阻碍膜内单分子运动. 考虑到 MeβCD 处理主要去除细胞膜脂筏内的胆固醇, 该现象可能是由于脂筏结构破坏后释放出的大量饱和脂质和膜蛋白阻碍膜内单分子扩散运动而导致的.

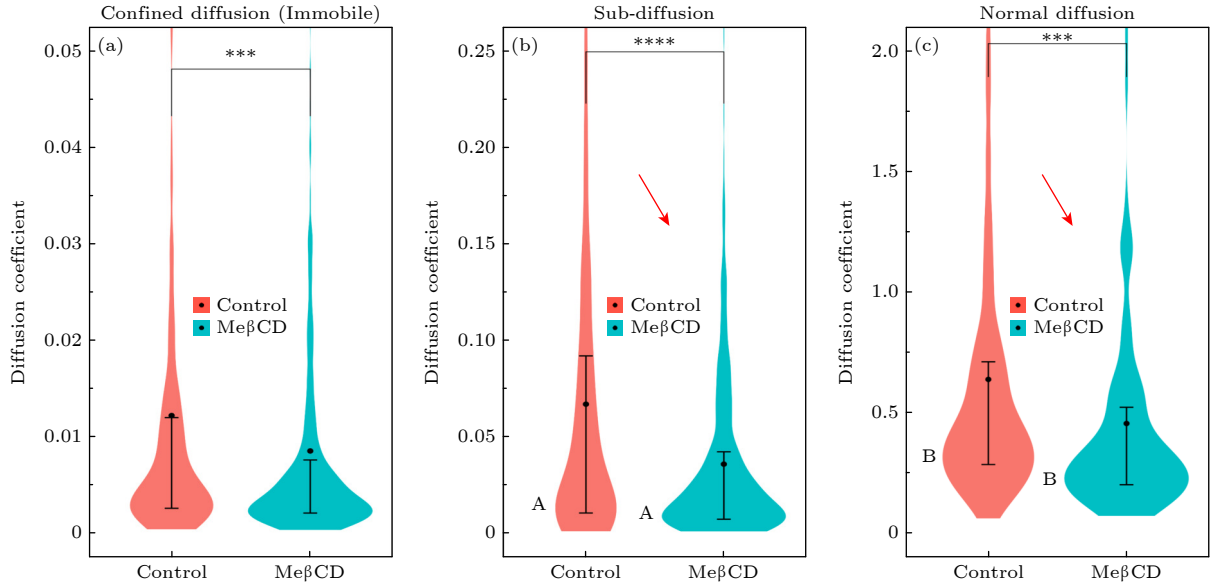


图 3 B16 活细胞膜 (Me β CD 处理前或处理后) 内单分子扩散运动分析, 即利用“两步归类法”将其分为受限扩散 (Confined diffusion)、亚扩散 (Sub-diffusion)、正常扩散 (Normal diffusion) 3 个子群, 分别计算其单分子轨迹的时间平均 MSD 并作出其分布, 轨迹数目见表 2

Fig. 3. Analysis of single-molecule diffusion in live cell membranes (before or after Me β CD treatment). This analysis employs a two-step classification method to divide the data into subgroups of confined diffusion, sub-diffusion, and normal diffusion. The time-averaged MSD is calculated for individual trajectories and its distribution is plotted. The number of trajectories analyzed is provided in Table 2.

为了更直观比较模型膜与活细胞膜内单分子扩散运动的差异, 我们绘制了模型膜 (以纯 DOPC 和 DOPC:Chol = 1:1 混合膜为例) 和活细胞膜 (Me β CD 去除胆固醇之前及之后) 内单分子扩散系数及异常指数的二维分布图 (图 4)、以及包括位移

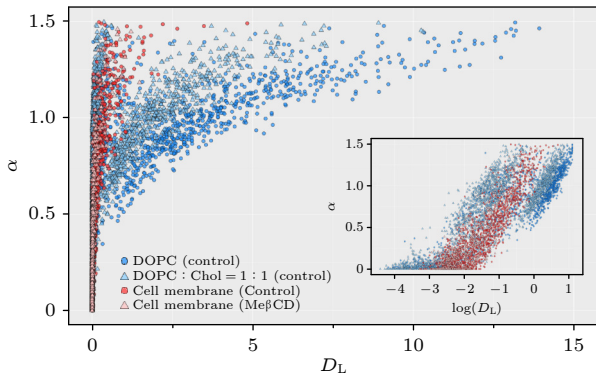


图 4 不同体系内 D_L 和 α 值分布, 包括纯 DOPC 膜 (包含 1324 条轨迹)、DOPC:Chol = 1:1 混合膜 (2347 条轨迹)、活细胞膜经 Me β CD 处理之前 (1436 条轨迹) 和之后 (784 条轨迹), 其中插图中横坐标以对数形式显示

Fig. 4. Distribution of D_L and α values in different systems: pure DOPC membrane (1324 trajectories), DOPC:Chol = 1:1 mixed membrane (2347 trajectories), living cell membrane before Me β CD treatment (1436 trajectories), and after Me β CD treatment (784 trajectories). The horizontal coordinate is presented logarithmically in the inset.

(displacement) 参数的三维分布图 (见补充材料模型膜和活细胞膜内单分子运动轨迹的扩散系数、异常指数、位移的三维分布视频 (online)). 由图 4 可以观察到 4 个体系的 D_L 值分布呈现出明显的差异: 纯 DOPC 膜的 D_L 分布明显高于 DOPC:Chol = 1:1 混合膜、以及活细胞膜和去除胆固醇后的活细胞膜; 同时, 它们都包含大量受限 (静止) 分子. 相较之下, 4 个体系的 α 值分布并无显著差别, 皆处于一个较大范围. 补充材料中的三维分布视频 (online) 进一步显示出位移 (即分子运动速率) 呈现与 D_L 类似的规律. 这些结果意味着, 在不同膜体系之间, 尽管单分子扩散运动速率存在较大差异, 但它们都是多种运动模式的混合, 都会涉及复杂的分子间相互作用.

5 结 论

针对细胞膜内单脂质分子扩散运动的复杂模式, 本文提出了一种无监督学习的“两步归类法”, 用于识别和分类单分子轨迹, 并将轨迹划分为受限扩散 (confined diffusion)、亚扩散 (sub-diffusion)、正常扩散 (normal diffusion) 以及超扩散 (super-diffusion) 等不同运动模式. 利用该方法, 实现了对

不同运动模式下分子轨迹子群的独立分析,并研究了 DOPC/Chol 组成的模型细胞膜和活细胞膜中单分子扩散运动行为特征以及胆固醇成分的影响. 结果显示,即使是单一成分的纯 DOPC 模型膜,其单脂质分子扩散运动也会呈现出不同的模式,反映出膜体系内复杂的分子间相互作用;此外,胆固醇的加入会降低脂质分子的扩散速率,增大受限分子和亚扩散分子的比例,且影响程度与胆固醇的比例正相关. 相较而言,活细胞膜内分子的扩散速率更慢、受限分子和亚扩散分子比例更高;并且,Me β CD 处理和去除胆固醇会进一步降低膜内分子扩散速率. 特别是,“两步归类法”为区分单分子运动轨迹提供了一种高效、定量的准确方法,且不受运动模式种类的限制;但不足之处在于,该方法以单条轨迹为基元开展分析,忽视了单条轨迹内不同时间阶段运动模式的可能差异. 单分子(单粒子)追踪技术在时间和空间的多尺度上提供了待测系统的动态变化信息,在生物及传统软物质研究领域具备独特优势和广阔应用前景. 本工作聚焦于模型及活细胞膜内的脂质分子运动,然而,该技术在膜蛋白激活和凝聚、蛋白质与细胞膜的相互作用、外源粒子入侵细胞等过程以及单细胞运动分析等研究中也都有着重要应用^[44–47]. 目前的单分子追踪主要关注分子在二维平面(如细胞膜内)的运动,这得益于较慢的分子扩散速率(比在细胞质中慢约 100 倍)、有限的 Z 方向上分子运动以及高选择性全内反射照明来降低背景噪声. 然而事实上,分子的跨膜运动以及膜曲率变化等也是备受关注的生物过程^[48]. 总的来说,本工作发展了区分单分子运动模式的方法,并基于此揭示了 DOPC 模型膜和活细胞膜内不同的单脂分子运动特征、以及胆固醇成分对两个体系的差异影响.

参考文献

- [1] Jacobson K, Liu P, Lagerholm B C 2019 *Cell* **177** 806
- [2] He W, Song H, Su Y, et al. 2016 *Nat. Commun.* **7** 11701
- [3] Golan Y, Sherman E 2017 *Nat. Commun.* **8** 15851
- [4] Subczynski W K, Pasenkiewicz-Gierula M, Widomska J, Mainali L, Raguz M 2017 *Cell Biochem. Biophys.* **75** 369
- [5] van Meer G, Voelker D R, Feigenson G W 2008 *Nat. Rev. Mol. Cell Biol.* **9** 112
- [6] Liu Y, Zheng X, Guan D, Jiang X, Hu G 2022 *ACS Nano* **16** 16054
- [7] Lyman E 2021 *Biophys. J.* **120** 1777
- [8] Zhang X, Barraza K M, Beauchamp J L 2018 *P. Natl. Acad. Sci. USA* **115** 3255
- [9] Chakraborty S, Doktorova M, Molugu T R, et al. 2020 *P. Natl. Acad. Sci. USA* **117** 21896
- [10] Pohnl M, Trollmann M F W, Bockmann R A 2023 *Nat. Commun.* **14** 8038
- [11] Fernandez-Perez E J, Sepulveda F J, Peters C, et al. 2018 *Front. Aging Neurosci.* **10** 226
- [12] Doole F T, Kumarage T, Ashkar R, Brown M F 2022 *J. Membr. Biol.* **255** 385
- [13] Byfield F J, Aranda-Espinoza H, Romanenko V G, Rothblat G H, Levitan I 2004 *Biophys. J.* **87** 3336
- [14] Yang S T, Kreutzberger A J B, Lee J, Kiessling V, Tamm L K 2016 *Chem. Phys. Lipids* **199** 136
- [15] Norregaard K, Metzler R, Ritter C M, Berg-Sorensen K, Oddershede L B 2017 *Chem. Rev.* **117** 4342
- [16] Ge F, Du Y, He Y 2022 *ACS Nano* **16** 5325
- [17] Chen P Y, Yue H, Zhai X B, Huang Z H, Ma G H, Wei W, Yan L T 2019 *Sci. Adv.* **5** eaaw3192
- [18] Jeon J H, Javanainen M, Martinez-Seara H, Metzler R, Vattulainen I 2016 *Phys. Rev. X* **6** 021006
- [19] Xu C, Yang K, Yuan B 2023 *J. Phys. Chem. Lett.* **14** 854
- [20] Xu C, Ma W, Wang K, He K, Chen Z, Liu J, Yang K, Yuan B 2020 *J. Phys. Chem. Lett.* **11** 4834
- [21] Pinholt H D, Bohr S S R, Iversen J F, Boomsma W, Hatzakis N S 2021 *P. Natl. Acad. Sci. USA* **118** e2104624118
- [22] Muñoz-Gil G, Garcia-March M A, Manzo C, Martín-Guerrero J D, Lewenstein M 2020 *New J. Phys.* **22** 013010
- [23] Cherstvy A G, Thapa S, Wagner C E, Metzler R 2019 *Soft Matter* **15** 2526
- [24] Granik N, Weiss L E, Nehme E, Levin M, Chein M, Perlson E, Roichman Y, Shechtman Y 2019 *Biophys. J.* **117** 185
- [25] Janczura J, Kowalek P, Loch-Olszewska H, Szwabinski J, Weron A 2020 *Phys. Rev. E* **102** 032402
- [26] Barkai E, Garini Y, Metzler R 2012 *Phys. Today* **65** 29
- [27] Krapf D, Metzler R 2019 *Phys. Today* **72** 48
- [28] Wu J F, Xu C, Ye Z F, Chen H B, Wang Y P, Yang K, Yuan B 2023 *Small* **19** 2301713
- [29] Yamamoto E, Akimoto T, Kalli A C, Yasuoka K, Sansom M S P 2017 *Sci. Adv.* **3** e1601871
- [30] Feder T J, Brust-Mascher I, Slatery J P, Baird B, Webb W W 1996 *Biophys. J.* **70** 2767
- [31] Briane V, Kervrann C, Vimond M 2018 *Phys. Rev. E* **97** 062121
- [32] Lanoiselée Y, Sikora G, Grzesiek A, Grebenkov D S, Wylomanska A 2018 *Phys. Rev. E* **98** 062139
- [33] Sikora G, Teuerle M, Wylomanska A, Grebenkov D 2017 *Phys. Rev. E* **96** 022132
- [34] Saxton M J, Jacobson K 1997 *Annu. Rev. Biophys. Biomol. Struct.* **26** 373
- [35] Kusumi A, Sako Y, Yamamoto M 1993 *Biophys. J.* **65** 2021
- [36] Saxton M J 1993 *Biophys. J.* **64** 1766
- [37] Shannon C E 1948 *Bell Syst. Tech. J.* **27** 379
- [38] Wright S J, Tenny M J 2004 *Siam J. Optim.* **14** 1074
- [39] Raja M A Z, Ahmed U, Zameer A, Kiani A K, Chaudhary N I 2019 *Neural. Comput. Appl.* **31** 447
- [40] Zhang Y, Yao F, Iu H H C, Fernando T, Wong K P 2013 *J. Mod. Power Syst. Clean Energy* **1** 231
- [41] Weron A, Janczura J, Boryczka E, Sungkaworn T, Calebiro D 2019 *Phys. Rev. E* **99** 042149
- [42] Hubicka K, Janczura J 2020 *Phys. Rev. E* **101** 022107
- [43] Xu R, Zhang W T, Jin T, Tu W, Xu C, Wei Y, Han W, Yang K, Yuan B 2024 *ACS Appl. Mater. Interfaces* **16** 6813
- [44] Li L, Ji J, Song F, Hu J 2023 *J. Mol. Biol.* **435** 167787
- [45] Li L, Hou R H, Shi X H, et al. 2024 *Commun. Phys.* **7** 174
- [46] Gao J, Hou R, Li L, Hu J 2021 *Front. Mol. Biosci.* **8** 811711
- [47] Lu Y, Ma J B, Teng C J, Lu Y, Li M, Xu C H 2018 *Acta Phys. Sin.* **67** 088201 (in Chinese) [陆越, 马建兵, 滕翠娟, 陆颖, 李明, 徐春华 2018 物理学报 **67** 088201]
- [48] Gao J, Hou R, Hu W, et al. 2024 *J. Phys. Chem. B* **128** 4735

COVER ARTICLE

Analysis of single-molecule diffusion movement in cell membrane based on unsupervised learning methods: Different effects of cholesterol on flowability of model membrane and living cell membrane*

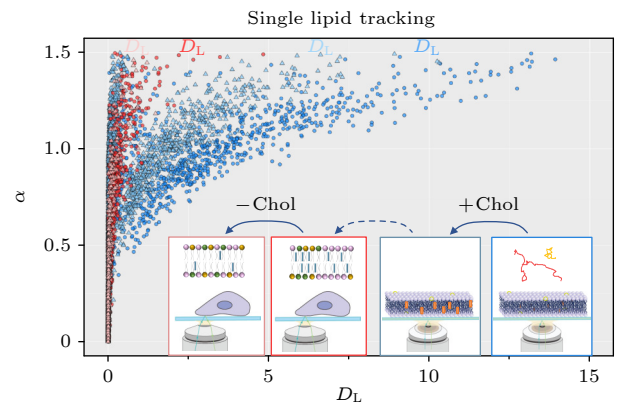
Tan Jin-Peng^{1)2)#} Zhang Wan-Ting^{2)3)#} Xu Cheng^{2)#} Lu Xue-Mei⁴⁾⁵⁾Zhu Wen-Sheng¹⁾ Yang Kai^{3)5)†} Yuan Bing^{2)‡}1) (*School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, China*)2) (*Songshan Lake Materials Laboratory, Dongguan 523808, China*)3) (*Center for Soft Condensed Matter Physics and Interdisciplinary Research, School of Physical Science and Technology, Soochow University, Suzhou 215006, China*)4) (*School of Optical and Electronic Information, Suzhou City University, Suzhou 215104, China*)5) (*Jiangsu Key Laboratory of Frontier Material Physics and Devices, Suzhou City University, Suzhou 215006, China*)

(Received 2 July 2024; revised manuscript received 31 July 2024)

Abstract

Single molecular tracking is a valuable approach to investigate the dynamic processes and molecular interactions in soft matter systems, particularly in biological systems. However, understanding the complexity of single molecule motion behaviors in biological systems remains a significant challenge. To address this issue, we propose a two-step classification method based on unsupervised learning to efficiently identify and classify single molecule trajectories. Firstly, we employ an entropy-constrained least square method to distinguish between confined (e.g., immobile) and unconfined diffusion trajectories. Subsequently, statistical tests are utilized to categorize the unconfined trajectories into different diffusion modes such as sub-diffusion, normal diffusion, and super-diffusion.

By applying this method, we analyze the diffusion motion of single molecules in both DOPC model cell membranes and living cell membranes while uncovering their distinct responses to cholesterol composition. Our findings demonstrate that both model membranes and living cell membranes exhibit diverse molecular diffusion modes. Specifically, in the DOPC model membrane system, the presence of cholesterol components



* Project supported by the National Natural Science Foundation of China (Grant Nos. 12274307, 32230063, 12347102, 22307090, 22203059), the Basic and Applied Basic Research Foundation of Guangdong Province, China (Grant No. 2023A1515011610), the Natural Science Foundation of Jiangsu Province, China (Grant No. BK20210100), and the Program of Jiangsu Key Laboratory of Frontier Material Physics and Devices, China (Grant No. KJS2131).

These authors contributed equally.

† Corresponding author. E-mail: yangkai@suda.edu.cn

‡ Corresponding author. E-mail: yuanbing@sslslab.org.cn

impedes lipid diffusion within the membrane. The degree of inhibition is positively correlated with the amount of cholesterol present. For instance, as the cholesterol content in the membrane increases from 0 to 20% (DOPC:Chol = 4:1) and 50% (DOPC:Chol = 1:1), there is an increase in the proportion of molecules, exhibiting confined diffusion and sub-diffusion (from 55% to 45%), while there is a decrease in the proportion of molecules, displaying normal diffusion and super-diffusion (from 45% to 35%). The ensemble diffusion coefficient of molecules in the membrane significantly decreases, which can be attributed to both a decrease in velocity among fast-moving molecules. Interestingly, after using Me β CD to remove cholesterol, the single-molecule mobility within the DOPC/Chol composite membrane system is restored to a level similar to that of the pure DOPC membrane.

Conversely, in the living cell membrane system, the diffusion coefficient values of molecules are significantly lower than those observed in the model membrane system; furthermore, the removal of cholesterol further slows down the molecular diffusion rate. This study contributes to understanding the intricacies of biomolecular motility and its dependence on environmental factors from a perspective of single molecular motion.

Keywords: single molecular motion, unsupervised learning, model cell membrane, cholesterol

PACS: 87.16.D-, 87.15.K-, 87.15.H-, 87.15.Vv

DOI: [10.7498/aps.73.20240915](https://doi.org/10.7498/aps.73.20240915)

基于无监督学习方法的细胞膜内单分子扩散运动分析：胆固醇对模型膜和活细胞膜流动性的不同影响

谭金鹏 张婉婷 徐成 卢雪梅 朱文圣 杨恺 元冰

**Analysis of single-molecule diffusion movement in cell membrane based on unsupervised learning methods:
Different effects of cholesterol on flowability of model membrane and living cell membrane**

Tan Jin-Peng Zhang Wan-Ting Xu Cheng Lu Xue-Mei Zhu Wen-Sheng Yang Kai Yuan Bing

引用信息 Citation: *Acta Physica Sinica*, 73, 188702 (2024) DOI: 10.7498/aps.73.20240915

在线阅读 View online: <https://doi.org/10.7498/aps.73.20240915>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

基于磁性隧道结的群体编码实现无监督聚类

Implementation of unsupervised clustering based on population coding of magnetic tunnel junctions

物理学报. 2022, 71(14): 148506 <https://doi.org/10.7498/aps.71.20220252>

红外及太赫兹辐照下细胞膜生物效应的研究进展

Research progress of biological effects of cell membrane under infrared and terahertz irradiation

物理学报. 2021, 70(24): 248707 <https://doi.org/10.7498/aps.70.20212030>

蜂毒肽与二元脂膜相互作用过程的单分子运动行为

Single molecular kinetics during the interactions between melittin and a bi-component lipid membrane

物理学报. 2020, 69(10): 108701 <https://doi.org/10.7498/aps.69.20200166>

蜂毒肽与单组分脂膜相互作用的单分子研究

Single-molecule study of interaction between melittin and one-component lipid membrane

物理学报. 2021, 70(17): 178701 <https://doi.org/10.7498/aps.70.20210477>

非监督学习高维多体波函数矢量轨迹所在低维子空间

Low-dimensional trajectory subspace of high-dimensional many-body wavefunctions by unsupervised learning

物理学报. 2021, 70(24): 247101 <https://doi.org/10.7498/aps.70.20210697>

基于单畴表征的高/低黏滞磷脂膜中的相分离

Phase separation in high/low viscosity phospholipid membranes based on single domain characterization

物理学报. 2022, 71(18): 188702 <https://doi.org/10.7498/aps.71.20220752>