

# 用荧光显微示踪方法研究 RecA 在 DNA 同源识别过程中的工作机理\*

庞哲 王爽 李辉 徐春华<sup>†</sup> 李明

(北京凝聚态物理国家实验室, 中国科学院物理研究所软物质物理重点实验室, 北京 100190)

(2012 年 4 月 15 日收到; 2012 年 5 月 23 日收到修改稿)

RecA 是原核生物体内参与 DNA 同源识别过程的一种关键蛋白, 长期以来一直是同源重组相关课题的重要研究对象. 通过荧光显微示踪方法, 发现在同源识别过程中 RecA 与单链 DNA 形成的核蛋白丝与模板 DNA 的结合是短时 ( $\tau = 0.2$  s) 和短程 ( $l = 1.05$   $\mu\text{m}$ ) 的, 结合后搜寻模板 DNA 上的同源位点的过程可分为布朗运动和定向运动两种模式. 结合时核蛋白丝并不是缠绕在模板 DNA 上, 而是以一种更弱的方式结合在模板 DNA 外侧进行位点搜寻. 如果在该过程中没有找到同源位点, 核蛋白丝就会脱离模板 DNA, 并寻找下一次与模板 DNA 结合的机会, 重复以上过程.

**关键词:** RecA, 同源识别, 单分子示踪, 液流拉伸

**PACS:** 87.15.H-, 82.37.Rs, 87.15.K-

## 1 引言

由于各种外源的物理或化学作用, 生物体细胞内的 DNA 经常容易产生缺损, 其中一种重要的缺损就是双链断裂. 双链断裂不仅会导致生物体的遗传信息缺失, 更有可能直接导致细胞的死亡. 为了最大程度地保证生物体的遗传信息不缺失并保护细胞, 生物体内有一套完整的响应机制, 对断裂的 DNA 进行修补. 最常见的一种应激机制就是重组修复. 在重组修复过程中, 涉及两条双链 DNA, 一条是断裂的双链 DNA, 另一条则是完整的模板 DNA. 重组修复过程一般分为四步. 第一步, 解旋酶与断裂的双链 DNA 平末端相结合, 并通过解旋作用产生一条 3' 端突出的单链末端. 在原核生物中, 这种酶一般是 RecBCD<sup>[1]</sup>, 真核生物中, 这种酶一般是以 Rad50 为核心的蛋白复合物<sup>[2]</sup>. 第二步, 大量的单链结合蛋白结合在这一单链末端保护单链 DNA, 并诱导一种同源重组酶规律地缠绕

在单链 DNA 上, 替换掉单链结合蛋白, 形成核蛋白丝. 在原核生物中, 这种酶一般是 RecA<sup>[3]</sup>, 真核生物中, 这种酶一般是 Rad51<sup>[4]</sup>. 第三步, 核蛋白丝在模板 DNA 上寻找到同源位点, 并使被缠绕的单链 DNA 替换掉模板 DNA 双链中与其完全一样的链, 形成三链结合体. 之后, 断裂 DNA 会以三链结合的位点为起点沿模板 DNA 进行延伸, 并最终恢复成为完整的双链 DNA. 最后, 特异性的酶将解开在三链复合体的位点留下的霍利迪 (Holliday) 交叉. 在原核生物中, 这种酶一般是 RuvABC<sup>[5]</sup>, 真核生物中, 这种酶一般是 Mus81<sup>[6]</sup>.

RecA 作为同源修复过程中非常核心的一种蛋白, 长时间以来一直受到科学界的关注<sup>[7-9]</sup>, 最近也有一些单分子的实验得到了一些阶段性的成果<sup>[8,9]</sup>. 目前达成的共识是 RecA 可以借助 ATP 或其类似物缠绕在单链 DNA 上形成核蛋白丝, 并进一步在双链 DNA 模板上搜寻到被缠绕单链的同源位点. 找到同源位点后, 核蛋白丝会沿单链 5' 至 3' 的方向逐渐缠绕到模板 DNA 上, 并将与其相同的

\* 国家自然科学基金 (批准号: 10927402, 11004234) 资助的课题.

<sup>†</sup> E-mail: xch@aphy.iphy.ac.cn

一根 DNA 置换出来,使核蛋白丝缠绕的单链 DNA 与模板 DNA 配对.最后,ATP 水解会促使 RecA 从单链的 5' 端开始逐渐解离下来.然而对于核蛋白丝如何快速地在大量的模板 DNA 序列中找到同源位点这一问题,科学界一直没有形成定论.最近 Forget 和 Kowalczykowski<sup>[10]</sup> 进行了一系列的实验,根据实验结果,他们猜测在模板 DNA 拉紧的状态下,核蛋白丝很难搜寻到同源位点;相反,模板 DNA 在松弛的状态下,核蛋白丝却可以通过一种奇特的方式最终与模板相结合.然而,结合过程终究只是他们的推测,核蛋白丝在模板上搜寻的细节与动力学参数也没有得到确认.我们的实验可能为解答这一问题提供了思路.

## 2 实验装置与材料

### 2.1 实验仪器

实验使用的观测仪器是我们借助 Olympus IX71 显微镜自行搭建的一套液流拉伸全内反射荧光装置.装置由 4 个部分组成.1) 显微镜主体,包括明场照明光路、目镜成像光路、CCD 成像光路,荧光滤光镜组使用 U-MWB2, U-MWG2 两组荧光滤光镜模块.2) 电荷耦合成像装置,也就是 CCD.我们使用的是 Andor iXon 885 EMCCD, CCD 芯片大小为  $1002 \times 1004$  像素,在装置中每像素对应的实际尺寸为  $160 \text{ nm}$ ,实验时增益参数为 300.在全屏录像模式下,使用的曝光时间为  $100 \text{ ms}$ .实验中,将每 4 个像素合并为 1 个像素,在全屏模式下可提高录像速度,曝光时间可提升至  $30 \text{ ms}$ .3) 激发光光路.实验中使用长春光机所生产制造的 CNI 牌半导体激光器(型号: MLL-III-473 150 mW, MLL-III-532 200 mW) 作为激发光光源,并自己搭建了全内反射光路.波长分别为  $473 \text{ nm}$  和  $532 \text{ nm}$  的激光从激光器发出,经过扩束器对光路进行调整后,通过反射镜与分光镜将两束光调为同轴,再被光束提升器提升至样品台以上高度,并以略大于全内反射角的某一固定角度入射,从而在一套装置中可以进行两种被激发物质的观测.在实验中,  $473 \text{ nm}$  的激光用来激发 YOYO-1,  $532 \text{ nm}$  的激光用来激发荧光小球.实验中,可以通过分别放置在光路中的两个电动快门,控制射入样品的激发光的波长.4) 样品池及棱镜.实验中我们自己设计了适用于该实验的

样品池装置.样品池由石英、夹层、盖玻片及样品流入/流出管组成.其中流入管接样品,流出管接一个可以对液体进行推送/抽取的流体泵.样品池的内表面通过处理可以标记链霉亲和素,方便将带有生物素标记的模板 DNA 固定在表面.通过自己制作的一种固定器将样品池固定在显微镜的载物台上,平放在显微镜载物台上后,棱镜被放置在石英片上方,将入射光导入样品池,即可实现全反射.整套装置如图 1 所示.

使用 NIH 的软件 ImageJ 对 CCD 记录的数据进行处理.原始数据首先被转化成 TIF 格式,导入 ImageJ 之后使用插件 ParticleTracker 捕捉荧光小球的位置信息.

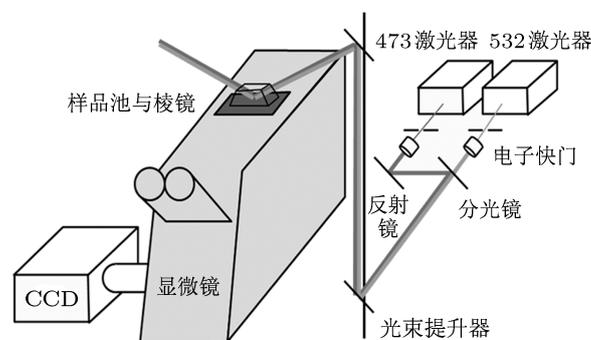


图 1 液流拉伸全内反射荧光装置示意图

### 2.2 实验材料

实验中使用的模板双链 DNA 为  $\lambda$  DNA (NEB-N3011L), 长度为  $48.5 \text{ kbp}$  (千碱基对). 使用两个带有生物素标记的  $12 \text{ nt}$  (核苷酸) 的单链将  $\lambda$  DNA 的末端进行标记, 可以制作出两个末端均标记有生物素的双链 DNA. 两个单链的序列分别为: GGCGGCGACCT-生物素和 AGGTCCGCCGCC-生物素. 制作核蛋白丝使用的单链 DNA 长度为  $90 \text{ nt}$ , 序列位于  $\lambda$  DNA 正向序列  $35.1\text{--}35.2 \text{ k}$  处, 具体序列为: 生物素-AAGAGCAATCTATGGTTTTGTGCGC-AGCCCTTAATGAAGGCAGGAAGTATGTGGTTA-CATCAAACAATTCCCATACATTAGTGAGTTGA. 由于生物体内断裂的双链 DNA 被解旋后形成一条  $3'$  突出的单链, 我们的单链也需要将  $5'$  端通过荧光小球封闭住, 因此在单链的  $5'$  位置带有生物素标记. 实验时将单链 DNA 稀释在 Tris-EDTA ( $10 \text{ mM}$  Tris,  $0.1 \text{ mM}$  EDTA,  $\text{pH} = 8.0$ ) 溶液中.

实验中使用的 RecA 是直接购买的蛋白 (NEB-M0249L), 分子量为 38.5 kD, 浓度为 2 mg/mL. 实验时使用 RecA 缓冲液进行稀释, RecA 缓冲液配方为加有  $\text{MgCl}_2$  的 Tris-HCl 溶液 (70 mM Tris-HCl, 10 mM  $\text{MgCl}_2$ , 5 mM DTT, pH = 7.60). 溶液配制完成后在  $-20^\circ\text{C}$  保存, 每次实验时取出将溶液回复至室温.

实验中对单链 DNA 进行荧光标记的荧光小球使用的是 Invitrogen 购买的 F8770, 直径为 40 nm, 结构为表面铺被有 neuroavidin 的聚苯乙烯小球, 小球内部填充了荧光素. F8770 的特征激发波长和发射波长为 580 nm/605 nm, 浓度大约为 600 nM. 实验中使用 YOYO-1 对模板 DNA 进行染色. 实验中使用的 YOYO-1 是从 Invitrogen 购买的 mp03600, 浓度为 1 mM. YOYO-1 是一种可以插在双链 DNA 碱基对中的荧光分子, 激发光和发射光的特征波长为 491 nm/509 nm.

实验时首先需要制备带有荧光小球标记的核蛋白丝. 取 2  $\mu\text{L}$  F8770 与 2  $\mu\text{L}$  单链 DNA 进行混合, 并放置在室温下过夜. 然后向混合物中加入 90  $\mu\text{L}$  RecA 缓冲液中, 并加入 10 mg/mL 的 ATP $\gamma$ S 1.6  $\mu\text{L}$ , 10 mg/mL 的 ADP 5.6  $\mu\text{L}$ . 在  $37^\circ\text{C}$  下反应 15 min, 样品制备完成后立刻使用. 为防止荧光小球在样品池表面的非特异性吸附, 使用前会在样品中加入 2  $\mu\text{L}$  浓度为 10 mg/mL 的 BSA 水溶液. 之所以使用的荧光小球浓度大大高于单链 DNA, 是为了能够尽量保证每个与模板 DNA 结合的核蛋白丝都带有荧光标记, 否则在后续实验中与带荧光标记的核蛋白丝很可能会形成竞争, 结合在模板上而不带来任何可记录的信号.

双链 DNA 的固定是通过液流拉伸的方法实现的. 使用 RecA 缓冲液将双链 DNA 配成浓度为 5 nM 的溶液, 并使用离心管分装为 100  $\mu\text{L}$  每管. 实验前将 DNA 样品连在样品池的流入管处, 使用流体泵将 DNA 样品抽入样品池中, 保证在样品流入管与流出管中有同样体积的 DNA 溶液. 然后使用流体泵以 20  $\mu\text{L}/\text{s}$  的速度进行反复推送/抽取. 10 min 后, 可见每直径约 100  $\mu\text{m}$  的视野中, 表面约有 3—5 根 DNA 沿水流方向被连好.

### 3 结果与讨论

#### 3.1 核蛋白丝在模板 DNA 上搜寻的过程是短时短程的

为了验证标记有荧光小球的核蛋白丝能否与模板 DNA 结合, 首先需要进行一个简单的实验. 将核蛋白丝稀释至 50 nM, 将两端带有生物素标记的  $\lambda$ DNA 稀释至 5 nM, 混合后于  $37^\circ\text{C}$  下反应 15 min. 将该样品使用离心管分装为 100  $\mu\text{L}$  每管, 采用上述液流拉伸的方法固定 DNA. 10 min 后, 向流体池内加入浓度为 2 nM 的 YOYO-1 溶液, 在 U-MWB2 二色镜组下, 可以看到 YOYO 标记的  $\lambda$ DNA. 更换到 U-MWG2 二色镜组, 则可以观察到荧光小球. 通过观察发现, 有大多数  $\lambda$ DNA 上已固定了核蛋白丝, 说明在模板 DNA 完全游离的状态下, 标记有荧光小球的核蛋白丝可以与模板结合. 图 2(a) 显示了观测到的一根固定了核蛋白丝的模板 DNA, 图片为两个二色镜组下的图像进行叠加.

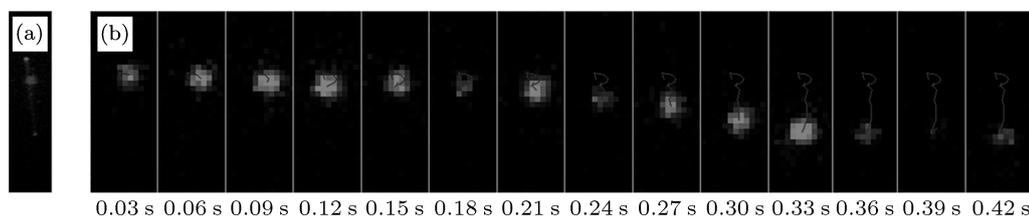


图 2 实验录像截图 (a) 将核蛋白丝与  $\lambda$ DNA 预先反应后, 再连入样品池中, 核蛋白丝被固定在模板 DNA 上, 图片为两组荧光信号叠加; (b) 先将  $\lambda$ DNA 连接在样品池中, 然后将核蛋白丝灌入样品池, 核蛋白丝沿模板 DNA 移动, 图片为荧光小球的信号

为了进一步观察到这一结合过程, 我们进行了以下实验. 先将两端带有标记的  $\lambda$ DNA 稀释至 5 nM, 并按照上述液流拉伸方法固定在样品槽

中, 并注入 2 nM 的 YOYO-1 溶液, 记录  $\lambda$ DNA 的位置, 再将一端标记有荧光小球的核蛋白丝稀释至 2 nM, 注入流体槽, 可观察到核蛋白丝在隐逝场

内做自由扩散. 实验时使用的曝光时间为 100 ms.

荧光小球直径为 40 nm, 扩散系数可以通过简单的斯托克斯关系进行估算:  $D = \frac{k_B T}{6\pi\eta a}$ , 其中,  $D$  为扩散系数,  $k_B$  为波尔兹曼常数,  $T$  为绝对温度,  $\eta$  为溶液的黏滞系数, 在室温下我们使用水的黏滞系数代替,  $a$  为荧光小球直径.

单链 DNA 长度为 90 nt, 缠绕上核蛋白丝之后长度约为 46 nm, 半径约 6 nm<sup>[11]</sup>, 且具有一定的刚性. 其扩散系数可使用圆棒状粒子的扩散系数进行计算:  $D = \frac{k_B T \ln(0.683L/b)}{3\pi\eta L}$ , 其中  $b$  为核蛋白丝的半径,  $L$  为核蛋白丝的长度.

经计算, 荧光小球的扩散系数约为  $10 \mu\text{m}^2/\text{s}$ , 而核蛋白丝的扩散系数约为  $35 \mu\text{m}^2/\text{s}$ . 由于带有荧光小球的核蛋白丝形状并不规则, 因此不能准确计算其扩散系数, 但可以估计其扩散系数小于一个荧光小球但大于一个直径 86 nm 的球体. 因此, 可粗略估计其扩散系数约为  $7.5 \mu\text{m}^2/\text{s}$ . 由此, 核蛋白丝停留在该区域内的时间可以按照布朗运动扩散的公式估算:  $\tau = \langle L^2 \rangle / 6D$ , 其中  $L$  为隐逝场穿透深度,  $D$  为估算的带有荧光小球标记的核蛋白丝的扩散系数. 隐逝场穿透深度非常浅, 约 100 nm, 核蛋白丝停留在隐逝场内的时间非常短暂, 远小于 1 ms. 因此在 100 ms 的时间分辨率下, 我们观察到的绝大多数信号是一闪而过的亮斑.

最初信号很难被观察到, 这很有可能是由于搜寻过程持续时间较为短暂, 在 10 帧/s 的时间精度下不容易分辨. 因此, 为了最大限度地提高时间分辨率, 我们牺牲了一定的空间精度将 CCD 像素做了合并, 从而将曝光时间提高到 30 ms. 在这样的时间分辨率下, 我们发现大约每 100 s 会有 1—2 个信号能够在隐逝场中持续 150 ms 以上. 这些信号中大部分出现了核蛋白丝在模板 DNA 位置附近停留, 或沿着模板 DNA 方向运动的情况. 由于核蛋白丝自身在隐逝波场中极难停留超过 150 ms, 因此这很可能是核蛋白丝在双链 DNA 上进行位点搜寻的表现. 图 2(b) 展示了一个标记有荧光小球的核蛋白丝沿 DNA 方向持续时间较长的运动过程. 图片中荧光小球的亮度忽明忽暗是由于模板 DNA 自身的涨落较大, 使得结合在其上的核蛋白丝在垂直于观察表面的方向也产生了涨落, 因此时而会远离表面.

通过对这一行为的持续时间进行统计, 我们发现核蛋白丝在识别过程中停留在双链 DNA 上

的时间很短, 经过高斯拟合, 得到的中心峰值仅为  $\tau = 0.20 \text{ s}$ , 如图 3(a) 所示. 相比于过去其他 DNA 结合蛋白的单分子实验结果<sup>[12,13]</sup>, 这一时间非常短, 在时间分辨率较低 (如 10 帧/s) 的情况下, 这一现象很容易被忽略. 这很可能是长久以来核蛋白丝识别过程没有被观察到的主要原因<sup>[14]</sup>.

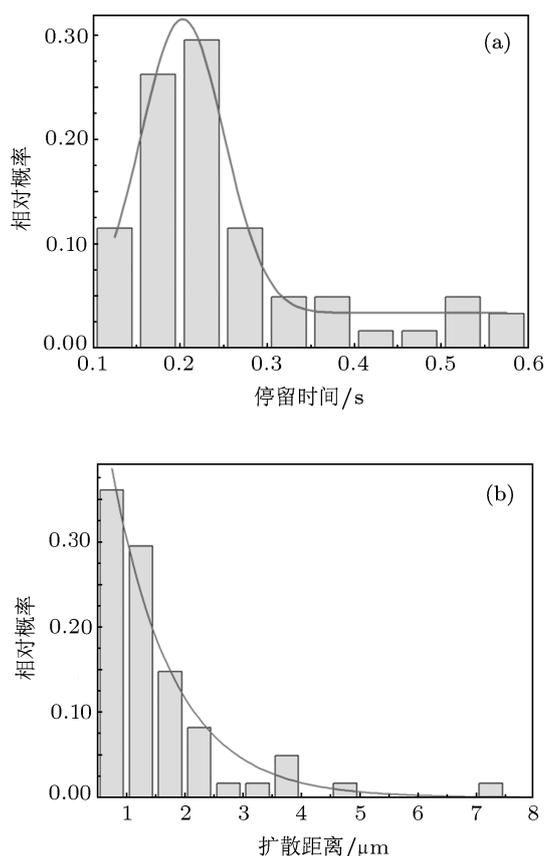


图 3 核蛋白丝在模板 DNA 上停留的时间与扩散距离的统计直方图 (a) 轨迹持续时间统计直方图; (b) 轨迹起始点与终点的距离统计直方图

我们对核蛋白丝在模板 DNA 上扩散的累计距离进行了统计分析, 经过指数拟合, 发现衰减常数仅为  $1.05 \mu\text{m}$ , 如图 3(b) 所示. 对于其他结合能力较强的蛋白而言,  $1 \mu\text{m}$  的扩散距离非常短<sup>[12,13,15]</sup>, 尤其是对于核蛋白丝这样具有复杂结构的大分子, 这说明核蛋白丝对于模板 DNA 的结合较弱.

### 3.2 在识别过程中, 核蛋白丝呈现两种扩散行为特征

由于核蛋白丝结合在模板 DNA 上时, 会不断外翻模板 DNA 的碱基进行检查, 因此核蛋白丝在

扩散过程中沿模板 DNA 方向经历的是忽高忽低的能垒, 扩散过程不一定是一个标准的一维布朗运动. 而我们观测到的数据持续时间一般较短, 无法直观地区分出某些曲线是否为布朗运动. 为了进一步对扩散行为的动力学特性进行区分, 我们采用了以下均方位移公式进行进一步数据处理:

$$\langle L^2 \rangle = 2D't^\alpha, \quad (1)$$

其中,  $t$  为选取的特定时间间隔,  $L$  为该时间间隔对应的空间位移,  $\alpha$  与  $D'$  同为待定参数. (1) 式常用来处理细胞膜嵌膜蛋白在细胞膜上的运动过程 [15,16]. 由于膜蛋白在细胞膜中扩散时也将遇到大小不一的能垒, 因此其扩散环境与我们的实验比较相似. 依据 (1) 式, 如果某次实验数据描述了一个布朗运动, 那么每一个时间间隔下的扩散行为都应该符合布朗运动的规律, 位移方均值与时间间隔的关系应为线性关系, 而如果某次实验数据描述的是定向运动, 那么位移方均值与时间间隔的关系则一定为二次关系. 因此在 (1) 式中, 当  $\alpha = 1$  时, 就是

典型的一维布朗运动,  $D'$  即为扩散系数; 当  $\alpha = 2$  时, 就是典型的一维定向运动,  $2D'$  即为扩散速度的平方. 而  $\alpha$  介于 1 和 2 之间时, 所描述的运动一般兼具定向移动和一维布朗运动的特性.

通过数据处理发现, 核蛋白丝在 DNA 上的扩散行为大致可以分为布朗运动 ( $\alpha_1 = 0.99$ ) 与定向运动 ( $\alpha_2 = 1.79$ ) 两种模式, 如图 4(a) 所示. 其中, 一维布朗运动扩散系数为  $D'_1 = (1.00 \pm 0.58) \mu\text{m}^2/\text{s}$ , 而定向运动速度为  $v = (2D'_2)^{1/2} = (4.76 \pm 1.90) \mu\text{m}/\text{s}$ . 图 4(b)、(c) 分别显示了布朗运动与定向运动的核蛋白丝运动的典型轨迹. 其中图 4(b) 所示轨迹  $\alpha$  为 1.17, 属于布朗运动, 图 4(c) 所示轨迹  $\alpha$  为 1.86, 属于一维定向运动. 图中黑色曲线表示核蛋白丝沿 DNA 方向的运动, 灰色曲线表示核蛋白丝垂直 DNA 拉伸方向的涨落. 可以明显看到, 通过液流拉伸的方法, DNA 内部的张力并不强, 因此在垂直 DNA 拉伸的方向上涨落可以达到  $0.3 \mu\text{m}$ .

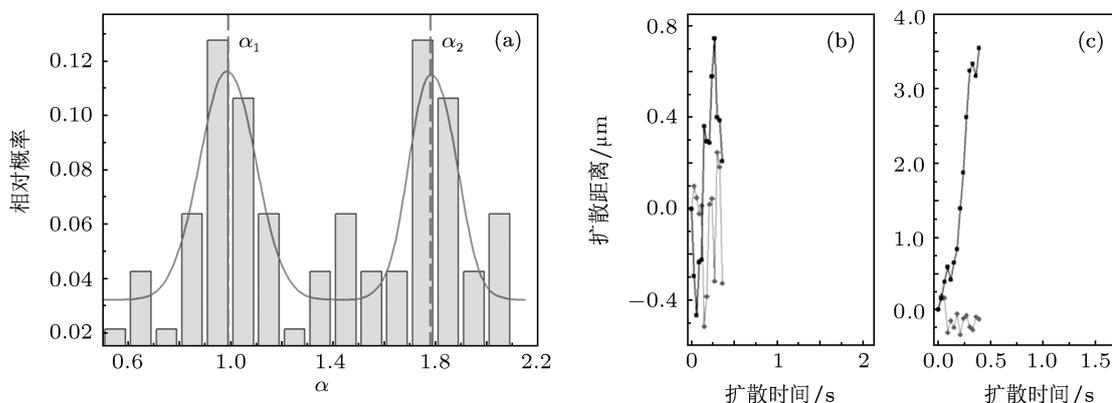


图 4 通过  $\alpha$  参数的拟合区分扩散行为的动力学特性 (a) 对  $\alpha$  参数的拟合; (b) 一维布朗运动典型轨迹; (c) 一维定向运动典型轨迹

### 3.3 讨论

长期以来围绕同源重组与同源修复的一个核心问题是: 修复蛋白缠绕单链 DNA 形成的核蛋白丝是通过怎样的动力学过程寻找到模板 DNA 上的同源位点的. 通过本文的实验, 我们大体可以确认, 核蛋白丝在模板 DNA 上寻找同源位点的搜寻行为如图 5 所示: 第一步, 核蛋白丝通过三维扩散结合到模板 DNA 上, 并通过一侧的结合位点与双链外侧进行较弱的结合; 第二步, 核蛋白丝沿双链

进行短程扩散并寻找位点尝试进行配对, 扩散过程包含了一维布朗运动与定向运动两种模式; 第三步, 如果核蛋白丝在模板 DNA 上没有寻找到合适的位点, 就离开模板 DNA 进行三维扩散, 并寻找下一次与模板 DNA 结合的机会; 一旦与模板 DNA 再次结合, 就会重复第一步至第三步的过程.

为了保证这一结论的正确性, 使用荧光显微镜示踪法对同源识别过程进行研究. 首先要确定荧光标记的方法不会显著影响实验物质的扩散性质. 我们选用荧光小球是由于其亮度好, 持续时间长, 便

于长时间使用高时间分辨率进行追踪. 且荧光小球直径与核蛋白丝的长度相仿, 计算表明其扩散系数也与核蛋白丝相差不大. 因此, 使用荧光小球对核蛋白丝进行标记, 对其扩散性质影响并不大.

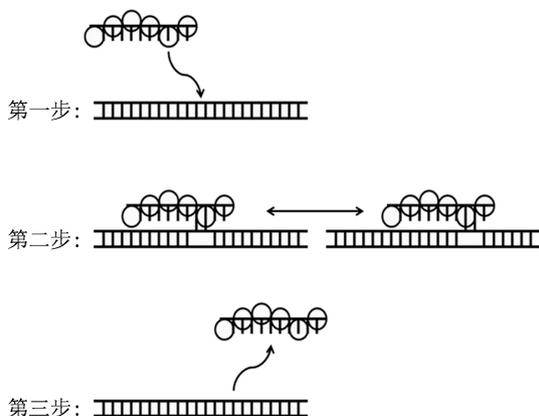


图5 核蛋白丝在模板 DNA 上搜寻同源位点的过程

我们每 100 s 只能观测到 1—2 次持续 150 ms 以上的信号, 这也是合理的. 按照 Debye-Smoluchowski 理论 (DS 理论), 许多学者已经对生物大分子在自由体系中碰撞 DNA 并结合的速率进行了计算<sup>[17]</sup>, 一般认为, 在 DNA 固定的情况下, 三维扩散系数为  $D_3$  的生物大分子, 如果其浓度为 1 M, 则在自由空间中碰撞到 1 bp DNA 并结合在其上的速率为:  $k_{DS} = 4\pi D_3 b$ , 其中  $b$  为 1 bp DNA 的长度, 即 0.34 nm. 按照实验中核蛋白丝三维扩散系数约为  $7.5 \mu\text{m}^2/\text{s}$  估算, 1 M 核蛋白丝通过三维布朗运动碰撞到 DNA 的速率约为  $10^8/\text{s}$ . 实验中为了降低荧光背景噪声, 增加软件对于荧光轨迹的识别率, 使用的核蛋白丝浓度约为 2 nM, 因此在 100 s 时间内, 约有 10 次左右的碰撞发生. 但 DS 理论是一个极端理想模型, 认为碰撞发生后生物分子一定会吸附在 DNA 上, 而实际中并非如此. 因此在实验中看到核蛋白丝以较低速率结合在 DNA 上, 并进一步搜寻同源位点是合理的.

同时, 应当说明的是, 整个系统中液流的冲击及机械漂移对实验数据几乎不造成影响. 在实验操作中, 我们非常注意液流对荧光小球的冲击, 当观察区域中有 DNA 存在并加入核蛋白丝后, 我们使用两个透析袋夹将样品池两端的进样管/出样管夹紧, 并等待数分钟后才进行录像, 在此情况下我们认为没有明显的溶液流动对实验数据造成影响. 同

时, 显微镜的载物台本身确实存在一定速度的不定向漂移, 但该速度非常缓慢, 大约在数 nm/s 的量级. 而实验中大多数现象持续时间远小于 1 s, 因此我们认为机械漂移对实验数据也不造成影响.

相关的磁镊实验数据表明<sup>[8]</sup>, 核蛋白丝缠绕在模板 DNA 上的速度为 2 nt/s, 而在我们的实验中, 核蛋白丝与模板 DNA 结合的总时间一般不超过 1 s, 同时扩散距离也非常短. 因此我们有理由认为, 核蛋白丝在搜寻过程中并没有缠绕在模板 DNA 上, 而是通过蛋白丝内部的某个结构域与模板 DNA 呈现一种弱结合<sup>[11]</sup>. 具体来说, 核蛋白丝通过在样品池中随机扩散, 会碰撞到模板 DNA, 并通过 RecA 的双链结合域结合在模板 DNA 的小沟中<sup>[18]</sup>. 但由于核蛋白丝是刚性结构, 因此与模板 DNA 结合时一定不是平行的, 而是沿小沟的螺旋切线方向, 与模板 DNA 的拉伸方向呈一定角度. 因此, 核蛋白丝最初与模板 DNA 的结合区域一定不大, 应该只是模板 DNA 磷酸骨架最外侧的 1 个碱基. 核蛋白丝与之碰撞后, 会通过翻转该碱基并与核蛋白丝缠绕的单链 DNA 中的碱基配对结合在模板 DNA 上. 所以在估算观察等待时间时选用 1 bp 作为碰撞区域的直径, 也是由核蛋白丝与模板 DNA 的这种结合方式决定的.

通过数据拟合发现, 核蛋白丝的运动模式大致可以分为布朗运动与定向运动两种模式. 这是因为核蛋白丝与模板 DNA 的结合方式非常复杂, 当核蛋白丝通过一个碱基结合在 DNA 上后, 为了行使功能, 会在模板 DNA 上继续搜寻位点. 可以通过对扩散系数的分析进一步判别核蛋白丝的搜寻模式是沿模板 DNA 的磷酸骨架进行搜寻, 还是沿着 DNA 的拉伸方向进行滑动.

核蛋白丝如果沿着模板 DNA 的磷酸骨架进行位点搜寻, 运动路径就限定在了模板 DNA 的小沟上, 在搜寻过程中, 核蛋白丝就会围绕模板 DNA 旋转. 这种情况下, 已有实验证实<sup>[19]</sup> 其扩散系数应为

$$D_h = b^2 \frac{k_B T}{[6\pi\eta R b^2 + 8\pi\eta R^3 + 6\pi\eta R (R_{OC})^2]} \times F(\epsilon), \quad (2)$$

其中,  $b$  为模板 DNA 的螺距, 一般认为是 3.4 nm;  $R_{OC}$  为在核蛋白丝与模板 DNA 配对的位置上, 核蛋白丝中心与模板 DNA 中心的距离, 一般认为是 7 nm;  $R$  为绕模板 DNA 旋转的生物大分子的半

径, 由于核蛋白丝标记了荧光小球, 荧光小球对扩散系数的影响更大, 因此在这里我们使用荧光小球的半径 20 nm 进行计算;  $F(\varepsilon)$  则表示核蛋白丝扩散过程中起伏不平的能垒对扩散系数的影响, 按照相关计算结果, 如果核蛋白丝在扩散过程中要从一个碱基滑动到另一个, 需要跨过的能垒高度的均方根为  $\varepsilon$ , 则  $F(\varepsilon) = \exp[-(\varepsilon/k_B T)^2]$ . 由于旋转方式进行搜寻的扩散系数与生物大分子半径的三次方呈倒数关系, 因此会比三维扩散系数小很多. 按照 (2) 式, 我们的实验中应当观测到沿模板 DNA 一维布朗运动的扩散系数约为  $0.07 \mu\text{m}^2/\text{s}$ , 比三维扩散系数要小两个数量级. 但在我们的实验中, 当  $\alpha = 1$  时得到的扩散系数为  $D'_1 = (1.00 \pm 0.58) \mu\text{m}^2/\text{s}$ , 虽然小于三维扩散系数, 但与上面的理论值相差比较大, 这说明核蛋白丝并不是靠旋转扩散进行位点搜寻的.

那么核蛋白丝就很有可能是沿模板 DNA 拉伸的方向进行滑行. 在这种情况下, 显然对扩散系数有影响的只有  $F(\varepsilon)$ . 核蛋白丝在扩散过程中遇到的起伏不平的能垒主要来自于核蛋白丝滑动时模板 DNA 上不同的碱基与 RecA 之间的相互作用. 因此, 我们可以得到一维滑动扩散的扩散系数  $D_s$ :

$$D_s = \frac{k_B T}{6\pi\eta a} F(\varepsilon) = D_3 F(\varepsilon), \quad (3)$$

其中  $D_3$  为三维扩散系数. 有实验数据表明, 一些与 DNA 修复有关的蛋白在沿模板 DNA 扩散时遇到的能垒起伏大约有  $1k_B T$ <sup>[20]</sup>. 因此可以得到滑动扩散的扩散系数约为  $2.7 \mu\text{m}^2/\text{s}$ , 这与我们观测到的扩散系数差别不大. 但观察到的扩散系数仍然比较小, 我们认为有三点原因: 首先, 由于液流拉伸方法固定的  $\lambda$ DNA 两端拉伸较为松散, 内部张力不足, 因此  $\lambda$ DNA 沿自身方向涨落较大, 空间位置涨落在平行和垂直于拉伸的两个方向均达到了  $0.3 \mu\text{m}$ , 而扩散过程持续时间又很短, 因此容易造成较大实验误差; 第二, 我们对于  $D_3$  的估算并不准确, 由于核蛋白丝标记了荧光小球之后形状不规则, 难以准确计算扩散系数, 因此对  $D_s$  的估算也产生了影响; 第三, 我们对  $\varepsilon$  的估计使用的是  $1k_B T$ , 可能较真实情况偏低, 按照实验中的数据, 可以得到扩散过程中的能垒起伏约为  $1.27k_B T$ , 这也是合理的. 我们在实验中使用的是 ATP- $\gamma$ S, 并不具有水解功能, 因此核蛋白丝在位点搜寻过程中无法消耗 ATP, 整个扩散过程只能由热涨落驱动. 根据以

上分析, 热涨落依然可以驱动核蛋白丝进行滑行.

在  $\alpha$  接近 2 的区域, 我们认为核蛋白丝呈现出定向运动的行为, 这可能与核蛋白丝与模板 DNA 通过多个碱基配对后会定向缠绕的特性有关. 过去的实验表明, 在体外环境中, 核蛋白丝通过 RecA 结合到模板 DNA 上后, 在配对大于 8 个碱基的位点会停留下来<sup>[21]</sup>, 并沿单链 DNA 5'—3' 的方向逐渐缠绕在模板 DNA 上<sup>[22]</sup>. 在这个过程中, 模板 DNA 的碱基间距会发生约 1.5 倍的变化<sup>[23]</sup>. 我们推测, 核蛋白丝中的单链如果只与模板 DNA 中的一个碱基配对, 沿模板 DNA 序列的上下游方向的能垒起伏约在  $1k_B T$  左右, 但如果有多碱基配对, 由于碱基间距和螺距都会发生变化, 因此所造成的能垒就会大于单个碱基外翻. 同时由于相关实验表明, 核蛋白丝的缠绕方向具有定向性, 因此我们也有理由怀疑这一能垒起伏沿单链 DNA 5'—3' 的方向倾斜. 因此我们认为, 在滑动过程中, 如果核蛋白丝与模板 DNA 有数个碱基配对, 有可能导致核蛋白丝沿模板 DNA 的某个方向能垒逐渐下降, 从而引起热涨落可以推动的定向运动. 但这只是猜测的一种机制, 具体机制还有待其他更加细致的实验证明.

在这一区域中, 滑动的速度可以用简单的方法进行估算:

$$\langle L^2 \rangle = 2Dt = v^2 t^2, \quad (4)$$

从而, 在某个时间间隔  $t$  内观察到的平均速度  $v = \sqrt{2D/t}$ . 由于在定向运动的持续时间一般也不超过 0.2 s, 因此我们可以得到  $v = 3.2 \mu\text{m}/\text{s}$ , 与实验中观测到的  $v = (4.76 \pm 1.90) \mu\text{m}/\text{s}$  相差不大. 因此我们认为以上的解释也是合理的.

核蛋白丝在模板 DNA 上寻找同源位点时, 可能会处于以上两种方式的任一种, 因此在每次的观测录像中, 有可能观察到两种方式相结合的状态, 这就造成了有较大的概率统计出  $\alpha$  取值在 1 和 2 之间.

我们认为核蛋白丝与模板 DNA 的结合是短程和短时的, 是与核蛋白丝能够更加有效地寻找到同源位点有关. 虽然核蛋白丝在搜寻过程中呈现出了定向运动的性质, 但很难保证在某次搜寻过程中核蛋白丝始终保持定向运动的方式进行位点搜寻. 因此, 如果核蛋白丝只是通过在模板 DNA 上的一维布朗运动进行位点搜寻, 那么可以得到核蛋白丝遍历整个模板 DNA 的时间:  $t = \langle L^2 \rangle / 2D$ . 大肠杆

菌 DNA 长度为 1000  $\mu\text{m}$  左右, 按照核蛋白丝扩散系数计算, 其遍历整个模板 DNA 所需时间为 5 天左右, 这显然大大超出了生物体本身进行快速的同源修复的要求. 因此可以说, 如果核蛋白丝只在模板 DNA 上进行一维扩散, 会有相当长的时间在无意义的重复搜寻. 显然通过一维扩散与三维扩散相结合的方式可以有效地解决这一问题. 已有相关计算表明<sup>[17]</sup>, 在这两种扩散相结合的方式下, 核蛋白丝遍历整个模板 DNA 所需要的时间与 DNA 长度正相关, 即

$$t \approx \frac{L}{\sqrt{D/\tau}}, \quad (5)$$

其中,  $L$  为模板 DNA 的长度,  $D$  为核蛋白丝在模板 DNA 上进行一维扩散的扩散系数,  $\tau$  为核蛋白丝停留在模板 DNA 上的特征时间. 按照我们实验中得到的数据, 核蛋白丝如果通过一维扩散与三维扩散相结合的方式在大肠杆菌 DNA 上寻找同源位点, 其遍历模板 DNA 所需的时间约为 7 分钟, 这已经大大低于仅仅进行一维扩散搜寻所需的时间. 因此, 通过三维扩散与一维扩散相结合的方式也是保证搜寻效率的一种机制.

## 4 结论

RecA 作为一种原核生物中常见的蛋白, 在同源修复过程中起到了核心作用. 长期以来的研究都未能解释 RecA 是如何帮助受损 DNA 的单链末端在模板 DNA 上快速搜寻到同源位点的. 我们借助荧光显微镜示踪的手段, 通过较高的时间分辨率, 发现了这一过程实际上是短时短程的, 我们认为这是长时间以来这一过程被忽视的重要原因. 通过这种直观的观察, 我们推断在搜寻过程中核蛋白丝会通过三维扩散寻找到模板 DNA 并与其结合, 然后通过 RecA 以一种较弱的结合方式与模板 DNA 相结合进行位点搜寻, 而非缠绕上去. 实际上核蛋白丝与模板 DNA 的结合方式非常复杂, 这决定了搜寻的模式也较为复杂, 但大体可以分为一维布朗运动和定向运动两种模式, 搜寻过程中经常是两种模式互相结合的. 当核蛋白丝没有找到同源位点时, 会从模板 DNA 上脱落, 并通过三维扩散继续寻找到模板 DNA 并与其结合, 重复在模板 DNA 上的搜寻过程.

- [1] Muskavitch K M T, Linn S 1981 *Enzymes* **14** 233
- [2] Maser R S, Monsen K J, Nelms B E, Petrini J H 1997 *Mol. Cell Biol.* **17** 6087
- [3] Benson F E, Stasiak A, West S C 1994 *EMBO. J.* **13** 5764
- [4] New J H, Sugiyama T, Zaitseva E, Kowalczykowski S C 1998 *Nature* **391** 407
- [5] Constantinou A, Davies A A, West S C 2001 *Cell* **104** 259
- [6] Heyer W, Ehmsen K T, Solinger J A 2003 *Trends Biochem. Sci.* **28** 548
- [7] Shan Q, Cox M M 1997 *J. Biol. Chem.* **272** 11063
- [8] Heijden T V D, Modesti M, Hage S, Kanaar R, Wyman C, Dekker C 2008 *Cell* **30** 530
- [9] Joo C, Mckinney S A, Nakamura M, Rasnik I, Myong S, Ha T 2006 *Cell* **126** 515
- [10] Forget A L, Kowalczykowski S C 2012 *Nature* **482** 423
- [11] Story R M, Weber I T, Steitz T A 1992 *Nature* **355** 318
- [12] Tafvizi A, Huang F, Fersht A R, Mirny L A, Oijen A M V 2011 *Proc. Natl. Acad. Sci.* **108** 563
- [13] Gorman J, Chowdhury A, Surtees J A, Shimada J, Reichman D R, Alani E, Greene E C 2007 *Cell* **28** 359
- [14] Saxton M J, Jacobson K 1997 *Annu. Rev. Biophys. Biomol. Struct.* **26** 373
- [15] Graneli A, Yeykal C C, Robertson R B, Greene E C 2006 *Proc. Natl. Acad. Sci.* **103** 1221
- [16] Li H, Duan Z W, Dou S X, Wang P Y 2012 *Acta Phys. Sin.* **61** 068701 (in Chinese) [李辉, 段兆文, 窦硕星, 王鹏业 2012 物理学报 **61** 068701]
- [17] Bruinsma R F 2002 *Physica A* **313** 211
- [18] Nishinaka T, Shinohara A, Ito Y, Yokoyama S, Shibata T 1998 *Proc. Natl. Acad. Sci.* **95** 11071
- [19] Bagchi B, Blainey P C, Xie X S 2008 *J. Phys. Chem. B* **112** 6282
- [20] Blainey P C, Luo G B, Kou S C, Mangel W F, Verdine G L, Bagchi B, Xie X S 2009 *Nature Struct. Mol. Biol.* **16** 1224
- [21] PHSieh P, Camerini-Otero C S, Camerini-Otero R D 1992 *Proc. Natl. Acad. Sci.* **89** 6492
- [22] Rosselli W, Stasiak A 1990 *J. Mol. Biol.* **216** 35
- [23] Stasiak A, Capua E D 1982 *Nature* **299** 185

# A study on the mechanism of RecA in homologous recognition by using single molecule fluorescence tracking\*

Pang Zhe Wang Shuang Li Hui Xu Chun-Hua<sup>†</sup> Li Ming

(Beijing National Laboratory for Condensed Matter Physics and CAS Key Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China)

(Received 15 April 2012; revised manuscript received 23 May 2012)

## Abstract

RecA plays an important role in homologous recognition in prokaryotes, and it has become a hot point in homologous recognition related research since its discovery. We establish an assay by combining total internal reflection fluorescence and flow stretching to visualize in real time the motion of single RecA-ssDNA filaments which are tagged with fluorescent beads. This enables us to study the interaction of RecA-ssDNA filaments with their templates in the homologous recognition process. It is found that the searching and binding is a short-time ( $\tau = 0.2$  s) and short-distance ( $l = 1.05$   $\mu\text{m}$ ) process. Two distinguished motion modes for the RecA-ssDNA filament are observed, a Brownian motion and a directed motion. The observations suggest a model that a RecA-ssDNA filament just interacts weakly with the template DNA before it binds firmly to the template DNA. If no homologous site is found in a searching process, the filament drops off the template and repeats the searching process again until it finally finds its target.

**Keywords:** RecA, homologous recognition, single molecule tracking, flow stretching

**PACS:** 87.15.H-, 82.37.Rs, 87.15.K-

---

\* Project supported by the National Natural Science Foundation of China (Grant Nos. 10927402, 11004234).

<sup>†</sup> E-mail: xch@aphy.iphy.ac.cn