What matters for lac repressor search *in vivo* —sliding, hopping, intersegment transfer, crowding on DNA or recognition?

模型：霍悦　结果：孙昌昱　结论与讨论：徐莹

ppt及展示：孙昌昱

哪个是乳糖阻遏子体内搜索过程中最重要的？

A. Sliding（滑行）

B. Hopping（跳跃）

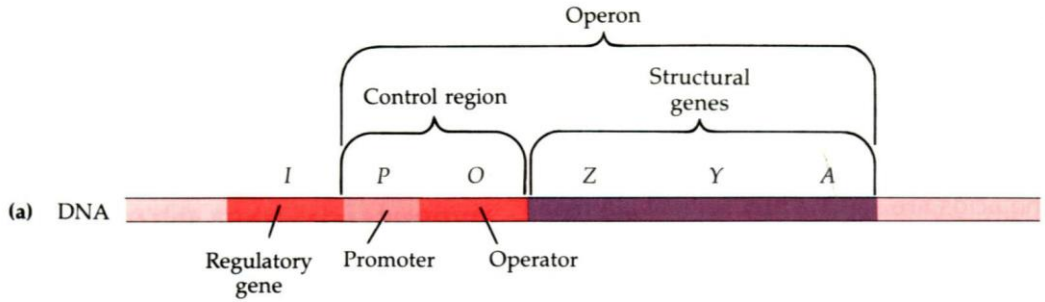
C. Intersegment transfer（段间转移）

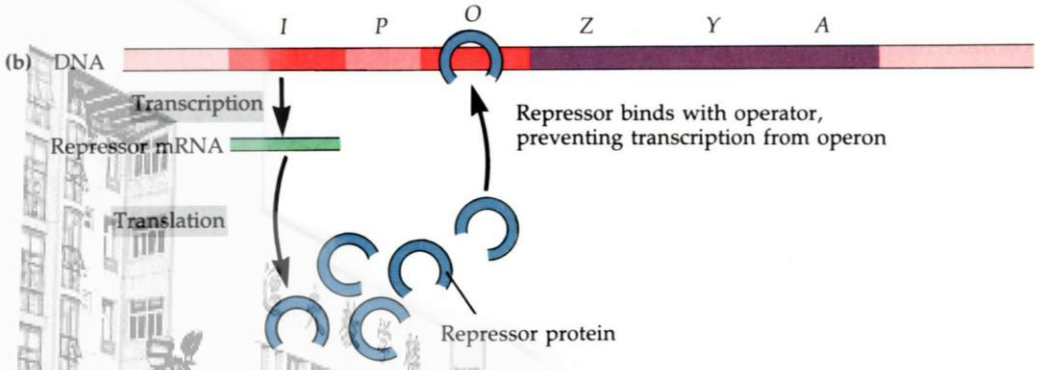
D. Crowding on DNA（DNA上堆积）

E. Recognition（识别）

- **A. Sliding（滑行）**
- 非特异性地结合在DNA上，然后顺着DNA滑行
- **B. Hopping（跳跃）**
- 非特异性结合的蛋白解离后又快速结合到原位点，极少数跳到旁边位点上
- **C. Intersegment transfer（片段间转移）**
- 蛋白转移到另一个不相关的远距离位点
- **D. Crowding on DNA（DNA上堆积）**
- DNA上本来就有蛋白质占据着，会影响其他蛋白结合
- **E. Recognition（识别）**
- Repressor达到操纵子位点时会发生识别

- 乳糖操纵子





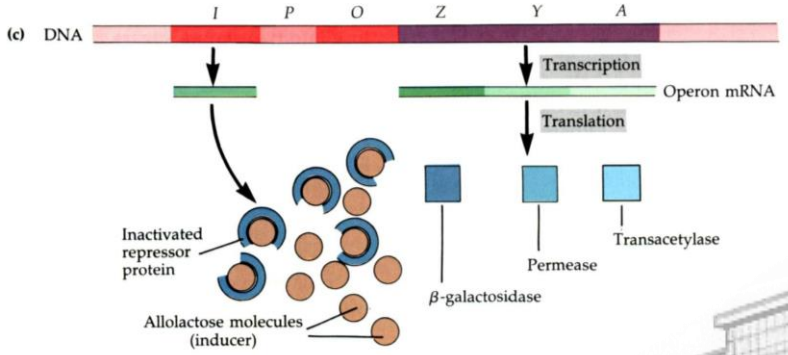


FIGURE 8.12 The *lac* operon. **(a)** This segment of the DNA molecule of *E. coli* shows the genes involved in lactose catabolism. In addition to the promoter and operator sites, the lactose operon contains three structural genes, which code for the proteins β-galactosidase, β-galactoside permease, and thiogalactoside transacetylase. For this operon the regulatory gene (*I*) is located next to the promoter (more often it is some distance away from the operon). **(b)** Control of the operon in the absence of lactose (repression). **(c)** Control of the operon in the presence of lactose (induction). When the inducer binds to the repressor protein, the repressor can no longer block transcription.

- We have investigated which aspects of transcription factor DNA interactions are most important to account for the recent *in vivo* search time measurements for the dimeric *lac* repressor.
- 我们研究了转录因子与DNA互作的哪些方面，对于最近的lac阻遏子二聚体的体内搜索时长，是最重要的。

- We find the best agreement for a sliding model where non-specific binding to DNA is improbable at first contact and the sliding LacI protein binds at high probability when reaching the specific O_{sym} operator.
- 我们认为滑动模型是最重要的：在初次接触时发生DNA的非特异性结合是不大可能的，而当滑动到特定的O_{sym}操纵子时，滑动的LacI蛋白则有很大的概率与DNA发生结合。

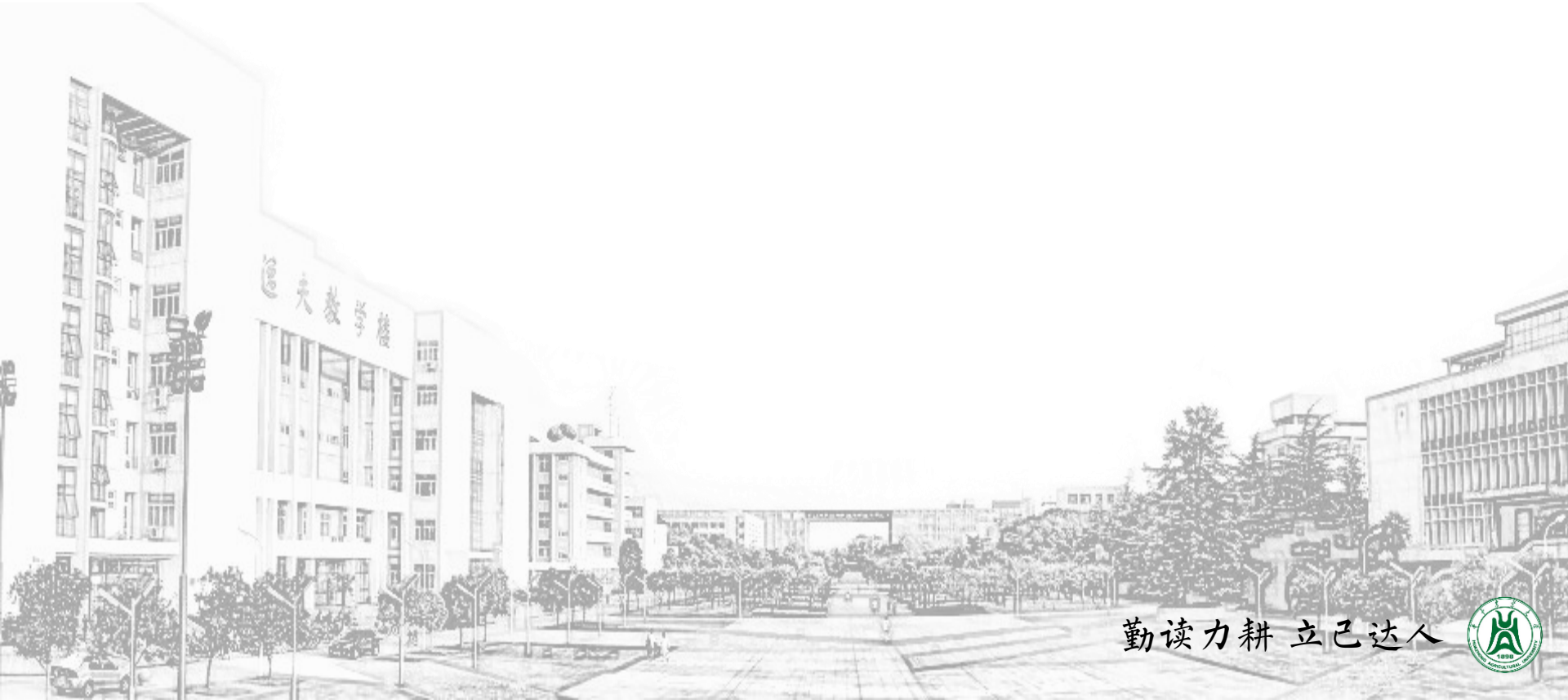
- We also find that the contribution of hopping to the overall search speed is negligible although physically unavoidable.
- 我们还发现跳跃对整体搜索速度的贡献可以忽略不计，尽管从物理理论的角度看是不可避免的。

- The parameters that give the best fit reveal sliding distances, including hopping, close to what has been proposed in the past, i.e. ~40 bp, but with an unexpectedly high 1D diffusion constant on non-specific DNA sequences.
- 拟合程度最好的参数揭示了滑行距离，包括跳行的，为大约40bp，和以前预测的相近，但在非特异性DNA序列上存在一个出乎意料的高一维扩散常数。

- Including a mechanism of inter-segment transfer between distant DNA segments does not bring down the 1D diffusion to the expected fraction of the *in vitro* value.
- 即便是加入了远距离DNA片段之间的片段间转移的机制，也并不能将一维扩散降低到体外检测值的预期波动范围内。

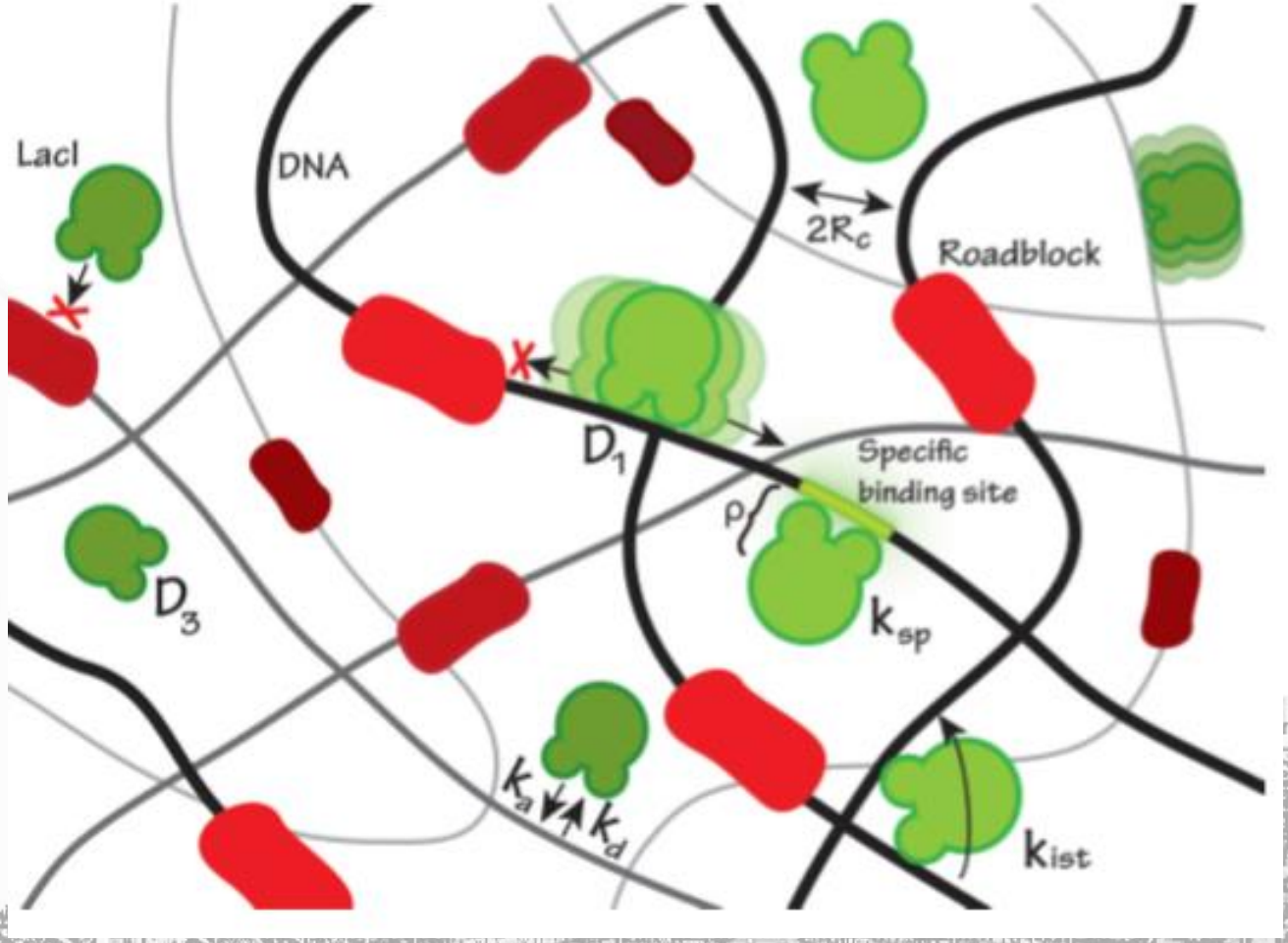
- This suggests a mechanism where transcription factors can slide less hindered *in vitro* than what is given by a simple viscosity scaling argument or that a modification of the model is needed.
- 这表明转录因子在体外的滑行阻碍，比简单的粘度缩放理论或该模型的修正版本给出的要少。

- For example, the estimated diffusion rate constant would be consistent with the expectation if parts of the chromosome, away from the operator site, were inaccessible for searching.
- 举个例子，如果对于染色体的部分而言，是处于操纵基因位点以外的话，搜索时蛋白质就无法进入，那么其预估的扩散速率常数与其期望值就应该是一致的。



1. **Introduction**
2. **Materials & Methods**
    1. The model
    2. The Monte Carlo simulation scheme
3. **Results**
    1. Approach
    2. Constraints from one or two stationary roadblocks
    3. The sliding length
    4. The effect of hopping
    5. Intersegment transfer
4. **Discussion**
    1. Parameter space—1D diffusion and sliding
    2. Parameter space—uncertainties
    3. Parameter space—crowding and the propensity to bind nonspecifically
    4. Parameter space—the recognition step
    5. Diffusion control and steric effects
    6. Hopping
    7. Intersegment transfer
5. **Conclusion**

- Sliding Model
- 假设：DNA是一个光滑的圆柱体，蛋白是一个完全反应球面



- Sliding Model

$$M_{\mathrm{acc}}\ell\pi R_c^2 = V_c.$$

- 计算DNA链间的体积

$$2\pi D_3\rho \left.\frac{\partial c}{\partial r}\right|_{r=\rho} = \kappa c(\rho),$$

- 描述了TF在DNA上解离又在一定范围内结合的动态过程

$$\phi_0 = \frac{\alpha \ln(R_c/\rho)}{1+\alpha \ln(R_c/\rho)}.$$

- 解离后结合到同一位点的概率

- Sliding Model $$\alpha = \frac{\kappa}{2\pi D_3} = \frac{k}{2\pi D_3 \ell}.$$
- 计算扩散控制程度（reactions occurs so quickly that the reaction rate is the rate of transport of the reactants through the reaction medium）

$$k_d = \lambda (1 - \phi_0) = \frac{\lambda}{1 + \alpha \ln (R_c/\rho)}.$$

- 非特异性结合下的解离速率

$$k_a = k(1 - \phi_0) = \frac{k}{1 + \alpha \ln (R_c/\rho)} = \frac{2\pi D_3 \ell}{1/\alpha + \ln (R_c/\rho)}.$$

- 非特异性结合下的结合速率

- Sliding Model

$$K_{\mathrm{RD}} = \frac{k}{\lambda} = \frac{k_a}{k_d} = \frac{F_B}{1 - F_B} \frac{V_c}{M_{\mathrm{acc}}}.$$

- 根据非特异性结合常数K_{RD}来计算出蛋白非特异性结合在DNA上的时间分数F_B

$$s = \sqrt{D_1 / k_d}.$$

- 蛋白在宏观非特异性结合时扫描的DNA距离

$$\tau = \frac{M_{\mathrm{acc}}}{2F_B\sqrt{D_1 k_d}}.$$

- 总的搜索时间

- Sliding Model

$$f_{\mathrm{sp}} = \left(1 + \frac{2\sqrt{D_1 k_d}}{k_{\mathrm{sp}}}\right).$$

- 由于存在一不小心“滑过头”了，没结合上的情况，所以加个修正指数

$$M_{\mathrm{acc}} = M_{\mathrm{tot}} v e^{1-\frac{1}{v}},$$

$$f_{\mathrm{cr}} = \frac{2s}{L(v)} = e^{1/v-1}\sqrt{1 + \frac{1-v}{vd}\sqrt{\pi \frac{D_1}{k_d}}}.$$

- 由于还有可能DNA上堆积了别的蛋白，所以再加个修正指数

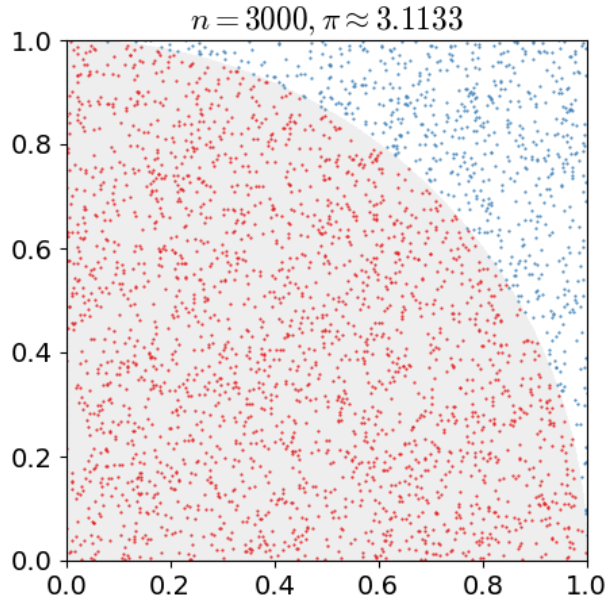
$$\tau = \frac{M_{\mathrm{tot}}}{2F_B}\sqrt{\frac{1}{D_1 k_d}\left[1 + \frac{1-v}{vd}\sqrt{\pi \frac{D_1}{k_d}}\right]}\left(1 + \frac{2\sqrt{D_1 k_d}}{k_{\mathrm{sp}}}\right)$$

- 最后修正过的总的搜索时间

- Monte Carlo simulation scheme（蒙特卡罗计算机模拟仿真）

- Monte Carlo methods (or Monte Carlo experiments) are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results.（Wikipedia）

- 蒙特卡洛方法（英语：Monte Carlo method），也称统计模拟方法，是1940年代中期由于科学技术的发展和电子计算机的发明，而提出的一种以概率统计理论为指导的数值计算方法。是指使用随机数（或更常见的伪随机数）来解决很多计算问题的方法。



- Monte Carlo simulation scheme（蒙特卡罗计算机模拟仿真）
- 由于有可能存在搜索区域内有多个operators，所以会出现roadblocks
- 用MC的方法来计算这种过程下的参数
- 首先随机一堆roadblocks旁边的蛋白质（是有分布的，与距离有关），然后迭代这个分布，直到将所有的roadblocks都模拟进去。
- 再模拟这些蛋白的行为：
- （i）以速率λ解离；
- （ii）以速率D_1向没有roadblocks的方向滑行1bp；
- （iii）如果有特异性结合位点，就以速率k_{sp}结合；
- （iv）以k_{IST}的速率进行片段间转移
- ↑Markov Process（马尔可夫过程）

- Monte Carlo simulation scheme（蒙特卡罗计算机模拟仿真）

$$p_{\text{dissoc}} = \frac{\lambda}{\lambda + 2D_1 + k_{\text{sp}} + k_{\text{IST}}},$$

- 解离的概率

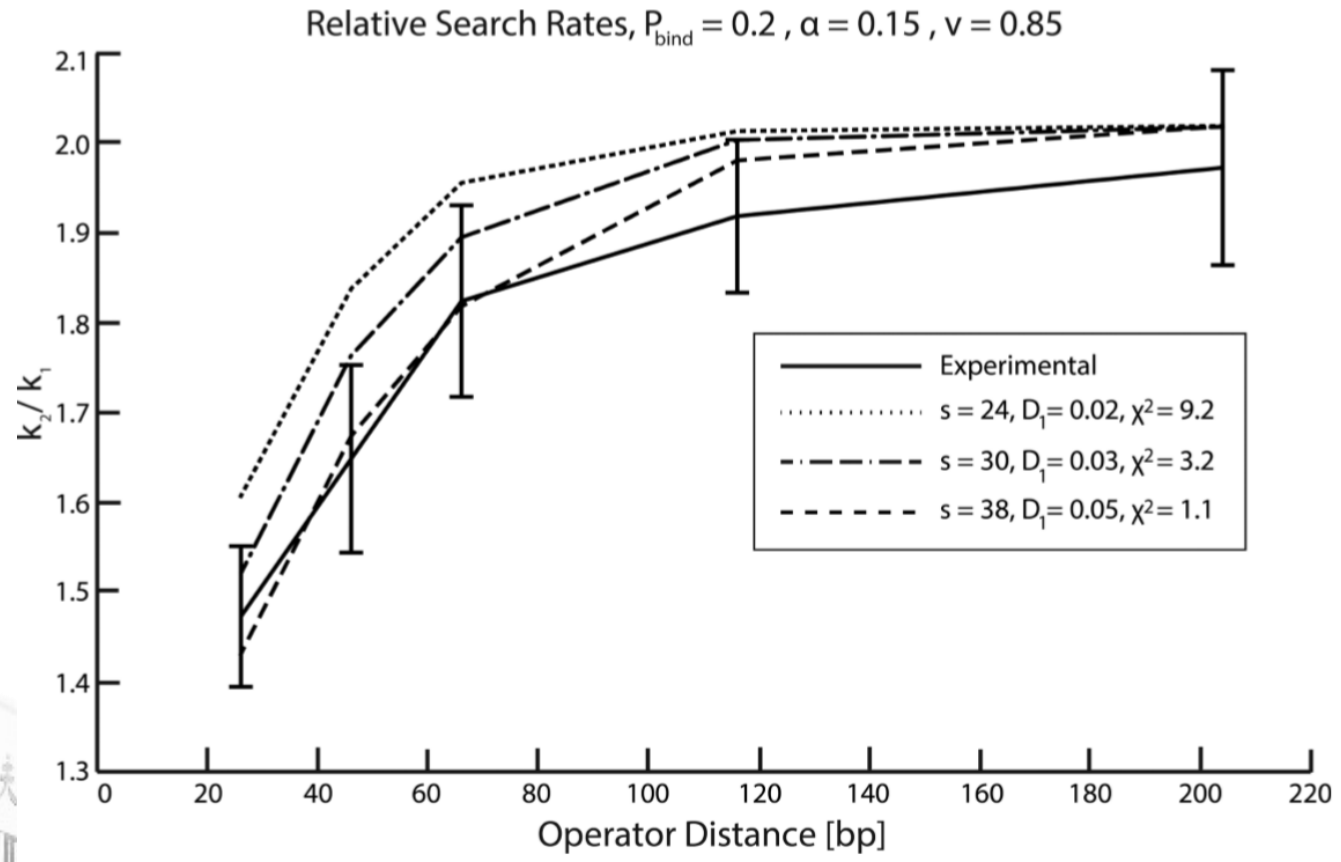
$$p_{\text{bind}} = \frac{k_{\text{sp}}}{\lambda + 2D_1 + k_{\text{sp}} + k_{\text{IST}}}.$$

- 结合的概率
- 重复个150 000次，来计算参数

$$\tau = \frac{N_{\text{micro}}}{\lambda F_B} = \frac{N_{\text{macro}}}{(1-\phi_0)\lambda F_B}.$$

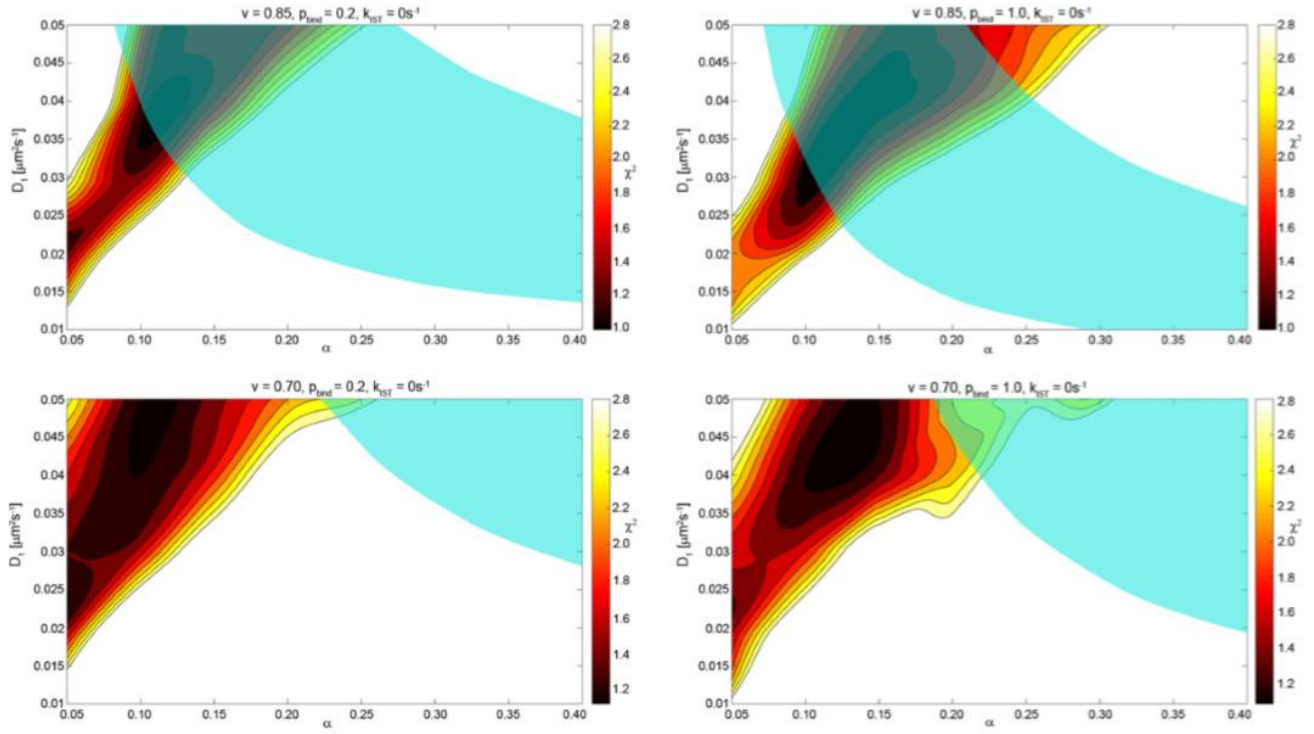
- 根据循环的结果得到总的搜索时间

- Results



- The figure shows examples of fits of simulations (dashed, dashdot and dotted curve) to experiments (solid curve) and the corresponding chi-square values used as constraints in determining the solution space.

- Results



- Red chi-square：good-fit of simulations to experiment
- Cyan region：absolute search time
- in the interval **236 s (three proteins)–416 s (five proteins)**
- Overlap：parameter space that can lead to agreement between experiments and the model

- Results

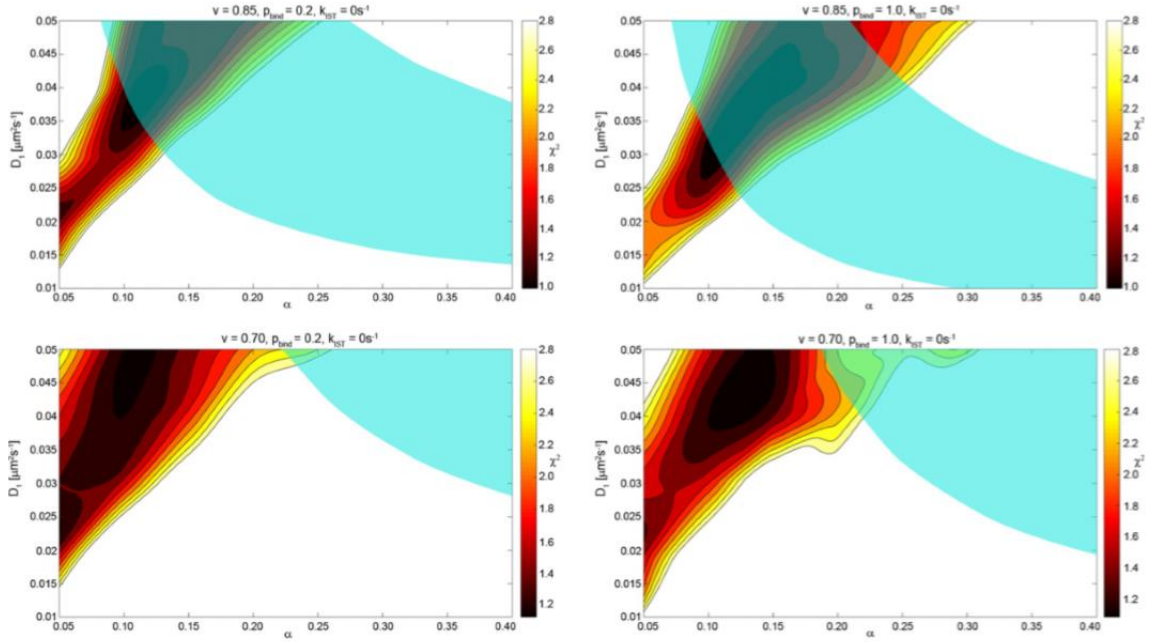
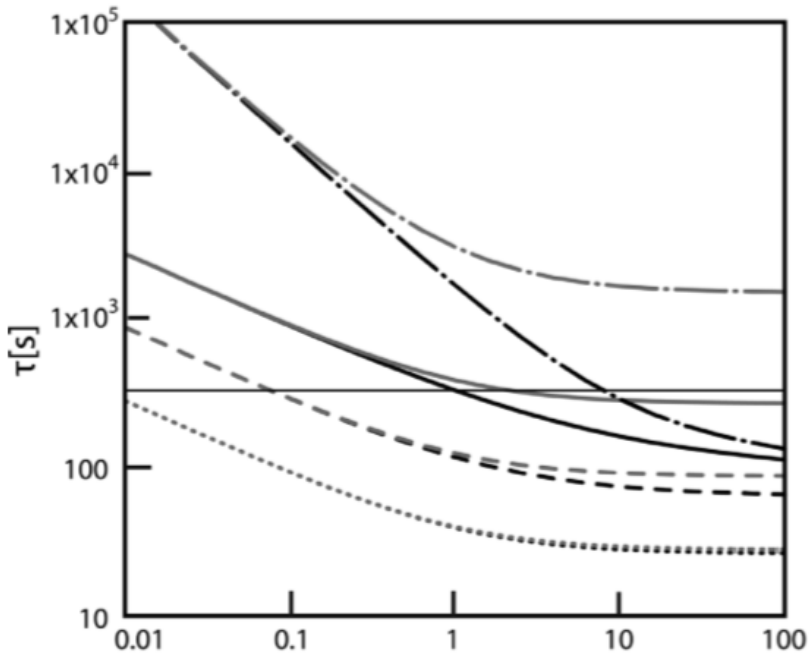


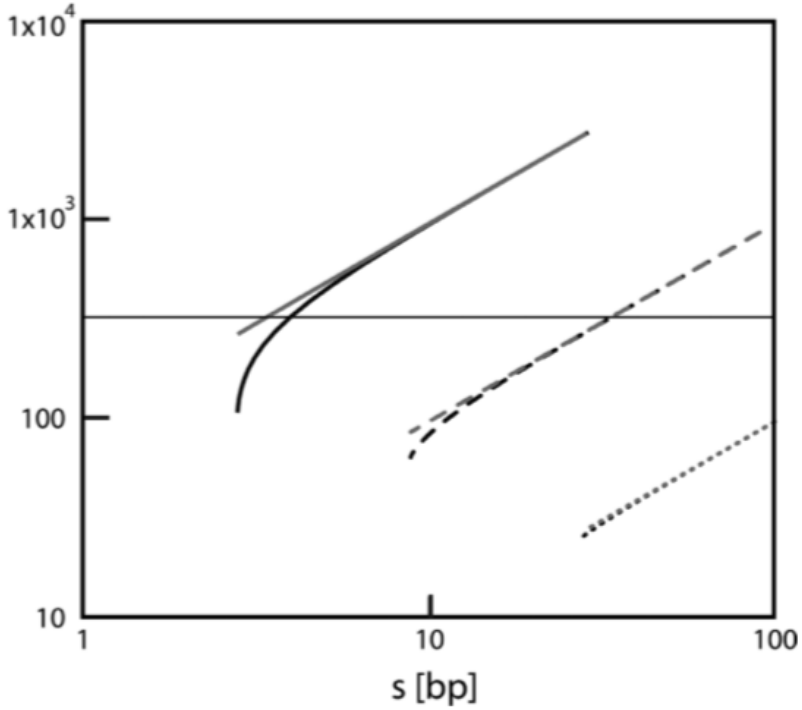
Table 2. Stationary roadblock next to the operator site, simulations and analytic results

	Simulations	Theory
($p_{\text{bind}} = 1.0,\ v = 0.85$)	1.66	1.70
($p_{\text{bind}} = 0.2,\ v = 0.85$)	1.60–1.50	1.63–1.55
($p_{\text{bind}} = 1.0,\ v = 0.70$)	1.35	1.40
($p_{\text{bind}} = 0.2,\ v = 0.70$)	1.33–1.26	1.37–1.32

- The experimental value is **1.75±0.18**

- Results

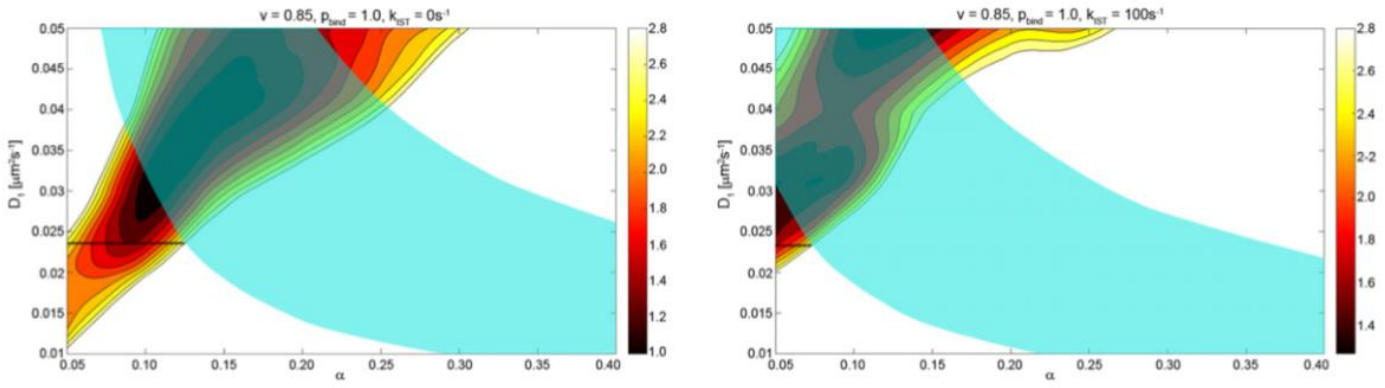




- Left panel: a function of α Right panel: a function of sliding length
- Black: with hopping Grey: without hopping

- Results

- 说是用O_{sym}做实验得到的单个蛋白的搜索时间是**81±2s**
- 和模型计算得到的不大符合，所以之前设的参数可能有问题，再重新改一改计算，然后没给结果（为啥呀）



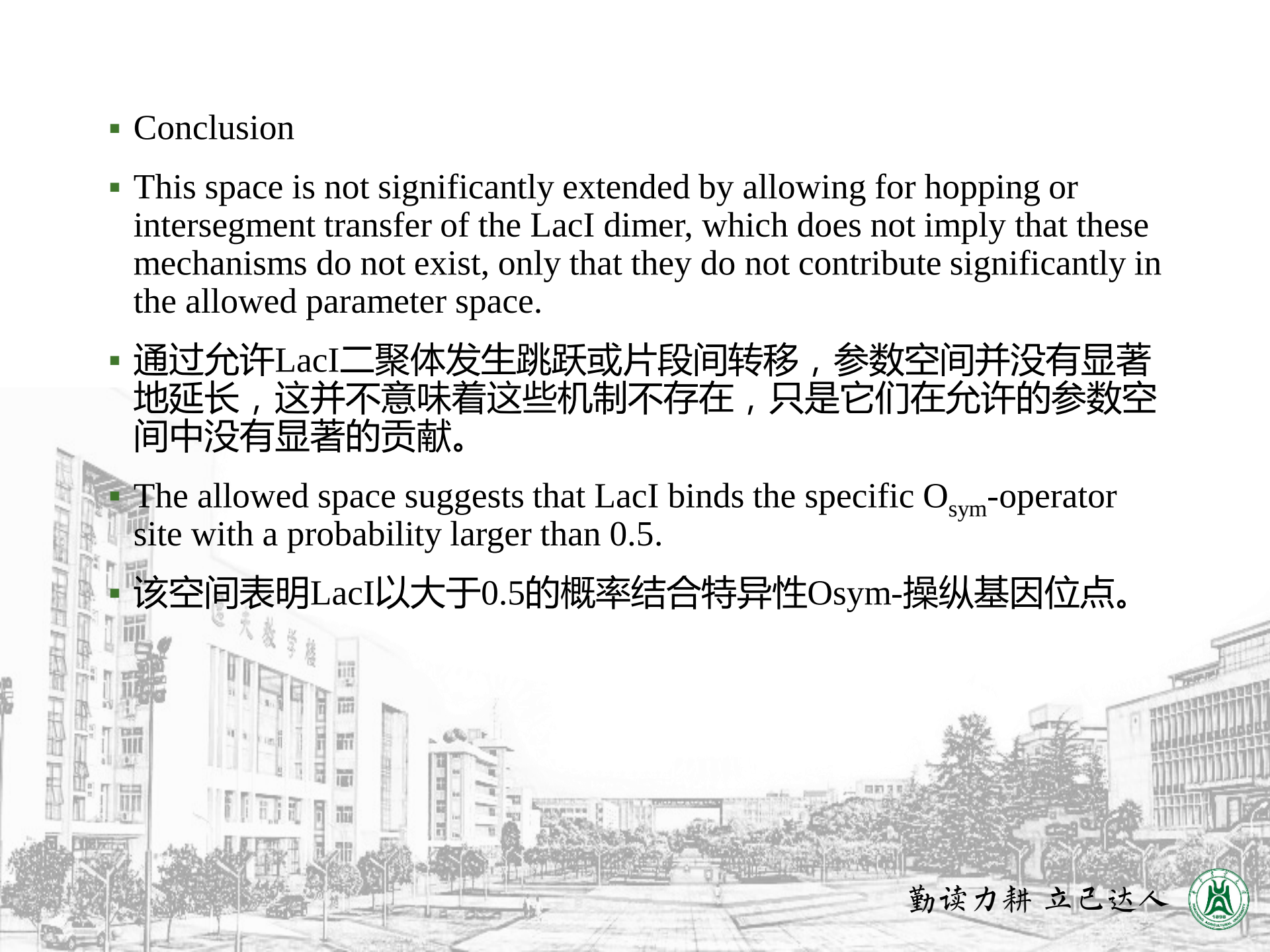
- The effect of **intersegment transfer** on the parameter space for the intersegmental transfer rate, k_{IST} = 100.
- 加上片段间转移，对结果影响也不大

- Discussion
- **结果有较大误差**
- 可能有两个原因：
- 1）模型或者参数错了
- 2）或体内搜索时一维滑行比体外的快（存在一维扩散）
- **不确定性**
- 最不确定的是同时有几个repressors在搜索，以及基因组的vacancy
- 结论：3~5个repressors同时搜索才行，多了就找不到位点了
- **蛋白堆积**
- 当F_B越趋向于1时，越容易发生特异性结合，越容易出现堆积
- **识别步骤**
- 滑动可以来回尝试结合，p_{bind}对这个速率有影响，但这个值很高，所以看不出来有什么改变（大于0.5时分辨不出来）

- Conclusion
- We have extended the sliding model to include crowding, hopping, intersegment transfer and the possibility of traversing the specific binding site to calculate the time it takes for one LacI molecule to find its specific binding site.
- 我们已经将滑动模型扩展到包括蛋白堆积，跳跃，片段间转移以及遍历特异性结合位点的可能性，以计算一个LacI分子找到其特异性结合位点所花费的时间。
- To account for recent *in vivo* data on the dependence between two operator sites as a function of the distance between them we also employed Monte Carlo simulations.
- 为了将最近的体内测量数据加入考量，我们还依照两个操纵基因之间的依赖关系作为它们之间距离的函数，采用了蒙特卡罗模拟。

- Conclusion
- Using the simulations and the analytical solutions as a constraint we generate solution spaces from a parameter sweep where the 1D diffusion coefficient and the degree of diffusion control were systematically varied given diverse values for the DNA occupancy, and the probability of LacI binding the specific operator site.
- 使用模拟和分析解决方案作为约束，我们从参数扫描中得到了解空间，其中一维扩散系数和扩散控制程度，在给定DNA占用的不同值，以及LacI结合特定操作者位点的概率时，会发生系统性地变化。
- We find that there exists a small parameter space where the model is compatible with the experimental measurements.
- 我们发现存在一个很小的参数空间，可以使模型与实验测量相兼容。

- Conclusion
- This space is not significantly extended by allowing for hopping or intersegment transfer of the LacI dimer, which does not imply that these mechanisms do not exist, only that they do not contribute significantly in the allowed parameter space.
- 通过允许LacI二聚体发生跳跃或片段间转移，参数空间并没有显著地延长，这并不意味着这些机制不存在，只是它们在允许的参数空间中没有显著的贡献。
- The allowed space suggests that LacI binds the specific O_{sym}-operator site with a probability larger than 0.5.
- 该空间表明LacI以大于0.5的概率结合特异性Osym-操纵基因位点。



- Conclusion
- We also establish that the lac repressor dimer binds non-specific DNA with a low probability at the first contact.
- 我们还确定了，在第一次接触时，lac阻遏物二聚体以低概率非特异性结合DNA。
- This does however not necessarily imply an energetic barrier for binding, but rather that not all patches on the repressor or DNA will bind at contact and that steric effects may need to be considered.
- 然而，这不一定意味着结合时存在能量障碍，而是并不是所有的阻遏物或DNA上的贴片都会在接触时结合，并且空间效应的影响可能是需要考虑进去的。

- Conclusion
- Finally, based on that the DNA occupancy by other proteins in the operator region seems to be ~15%, while the overall occupancy may be higher, we propose that chromosomal DNA may have two levels of coverage by Nucleoid Associated Proteins, one fraction that are randomly associated with DNA and thus are found in the operator region where they influence binding and sliding by the repressor and one fraction of more specifically binding proteins that do not reside in the operator region and only contribute to the search time by hiding part of the chromosome from searching.
- 最后，基于操纵基因区域中其他蛋白质的DNA占有率似乎是∽15％，而总体占用可能更高，我们提出了染色体DNA可能存在两种水平的NAPs覆盖率，一个部分是随机与DNA结合，因此存在于操作基因区域，在此它们通过阻遏物可以影响搜索蛋白的结合和滑动过程；另外一部分是更为特异性结合的蛋白质，它们不存在于操作基因区域，并且仅对通过染色体的隐藏部位的搜索时间起作用。

Fin

(finally)

勤读力耕 立已达人

- **请回答，2017**
- **Q1：到底哪个过程对于乳糖操纵子的搜索效率来说是最重要的？**
- A1：滑行（sliding）
- **Q2：用到了哪些建模的方法？**
- A2：动力学方程、蒙特卡罗模拟
- **Q3：描述一下repressor到底是怎么找到操纵基因的？**
- A3：先非特异性结合上，然后滑行一段距离s，滑不动了，蹦一下继续滑，如果中间遇到了roadblocks，那就片段间转移到另一个区域，直到找到自己的特异性结合位点，然后反复识别直到特异性结合上靶基因为止
- **Q4：到底为什么有这么大的误差？**
- A4：大概是模型出错了

- 搞不明白的点
- Q1：滑行的动力从哪儿来？
- Q2：为什么得到的结果与实验不符之后说是参数有问题，但又不给出换了参数计算得到的结果？
- Q3：在滑动模型的基础上加入跳跃和片段间转移的考量，然后分析得到说是影响很小，前提假设是不是就存在问题？
- Q4：参数的范围到底是怎么选定的？为什么后面给出的修改意见超出了给定的参数范围？
- Q5：反复计算更换参数后还是误差较大，为什么不考虑模型可能存在问题？
- Q6：如果认为误差是扩散导致的，那么这个扩散是否可以计算出来？（文章中提到用自由能，但没有给出具体的考虑方向）
- Q7：公式到底是怎么推出来的啊？补充材料里也没有讲清楚最前头几个公式和后面的是怎么联系在一起的。
- ……