Stochastic Model and Optimization of SELEX

Yue Wang

Department of Computational Medicine
University of California, Los Angeles

Joint work with Bhaven A. Mistry and Tom Chou

yuew@g.ucla.edu

Dec. 06, 2022

Reference

Yue Wang, Bhaven A. Mistry, and Tom Chou. (2022). “Discrete stochastic models of SELEX: aptamer capture probabilities and protocol optimization.” Journal of Chemical Physics, 156(24), 244103.

Outline

- Introduce SELEX: a process to select aptamers.
- Review the traditional deterministic model.
- Build a stochastic model for SELEX and analyze its properties.
- Search for the optimal protocol of SELEX.

Section I: Introduction

- Aptamers are short, single-stranded DNA or RNA molecules that bind to a specific target.
- Targets can be heavy metal ions, proteins, or even whole cells.
- Certain aptamers (linked with fluorescent tracers) can bind selectively to biomarkers on the cancer cells, but not to healthy cells. This test can identify cancer cells in a tissue sample.
- Besides testing, aptamers can also be used in treatment. Therefore, aptamers are also called chemical antibodies.

Section I: Introduction

- It is difficult to design and synthesize the best aptamer for a target directly.
- In general, we start with enough targets and a large library of randomly generated aptamers, and aptamers have different affinities to the target.
- How to select the best aptamers (with the highest affinities to the targets) in an easy way?

Section I: Introduction

- Systematic Evolution of Ligands by EXponential enrichment (SELEX): a convenient method to select the best aptamers.
- Aptamers with higher affinities to the targets are more likely to bind to the targets. We can use the targets to pick out such aptamers.
- It is similar to a population evolution process.

Aptamers and targets can bind and unbind reversibly.

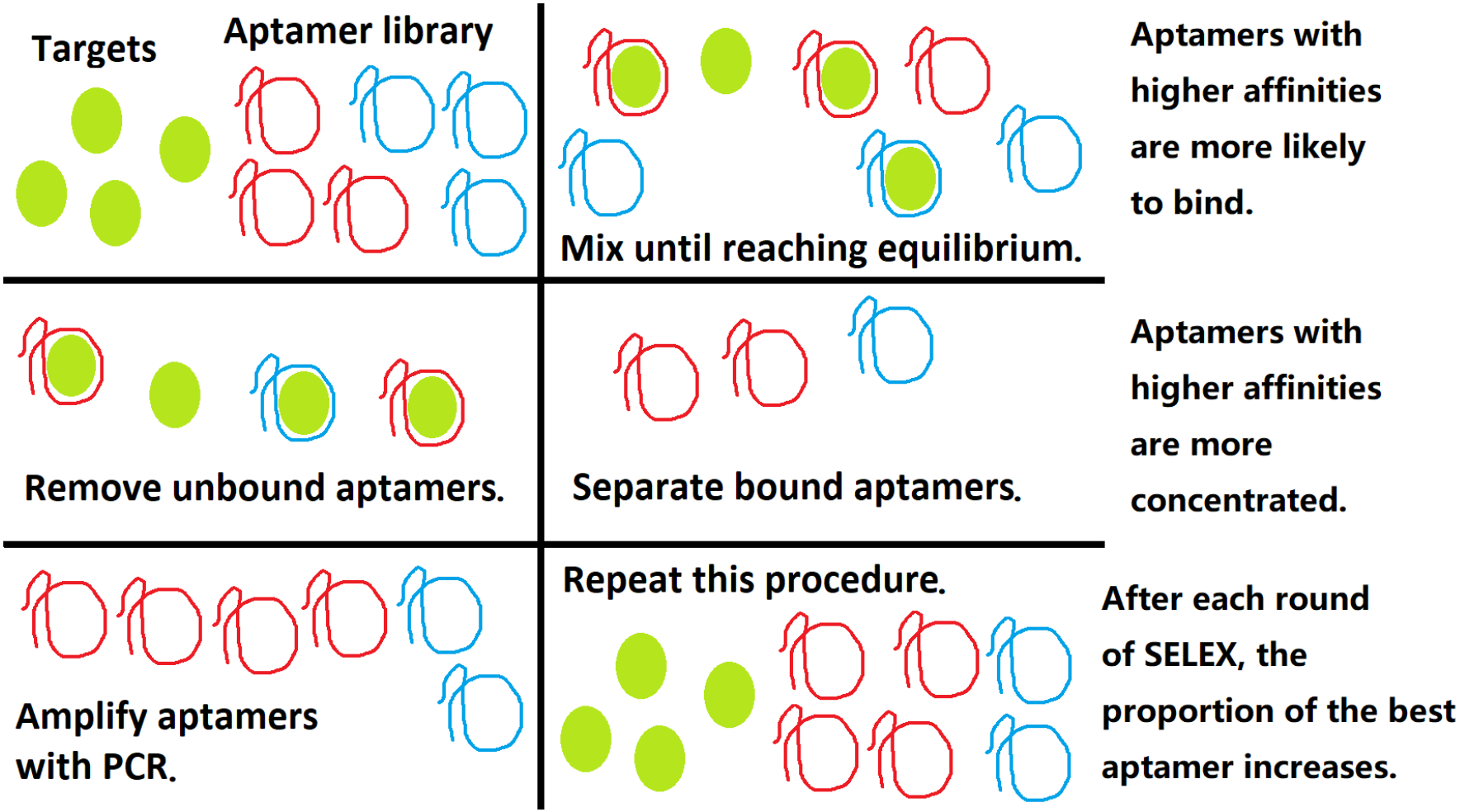


Figure: Protocol of SELEX

- We have enough targets, and the aptamers can be amplified by PCR. When starting one round of SELEX, we can control the quantity of targets and the quantity of aptamers, but the proportions of different aptamer types cannot be controlled.
- We obtain an optimization problem: maximize the proportion of the best aptamer (with the highest affinity) after this round of SELEX.
- A bad choice: add too many targets, so that almost all aptamers are bound. No selection is made, and the proportion of the best aptamer is invariant.

Section I: Introduction

- We need a mathematical model to study the optimization of SELEX.
- To simplify the discussion, we combine aptamers with different affinities into two types: strong type $\mathtt{A}_1$, weak type $\mathtt{A}_2$. The association constants (affinities) satisfy $K_1 > K_2$.

Section II: Deterministic Model

- A traditional deterministic approach uses the law of mass action, which is valid when the number of molecules is sufficiently large.
- Notations: $[S]$: total concentration of targets; $[A_i]$: total concentration of aptamer type $\mathtt{A}_i$; $[a_i]$: concentration of aptamers $\mathtt{A}_i$ that are bound to targets at equilibrium.
- At stationary, for each $i = 1, 2$ and the reaction $\mathtt{S} + \mathtt{A}_i \rightleftharpoons \mathtt{SA}_i$, we have:

$$([S] - [a_1] - [a_2])\,([A_i] - [a_i])\,K_i = [a_i].$$

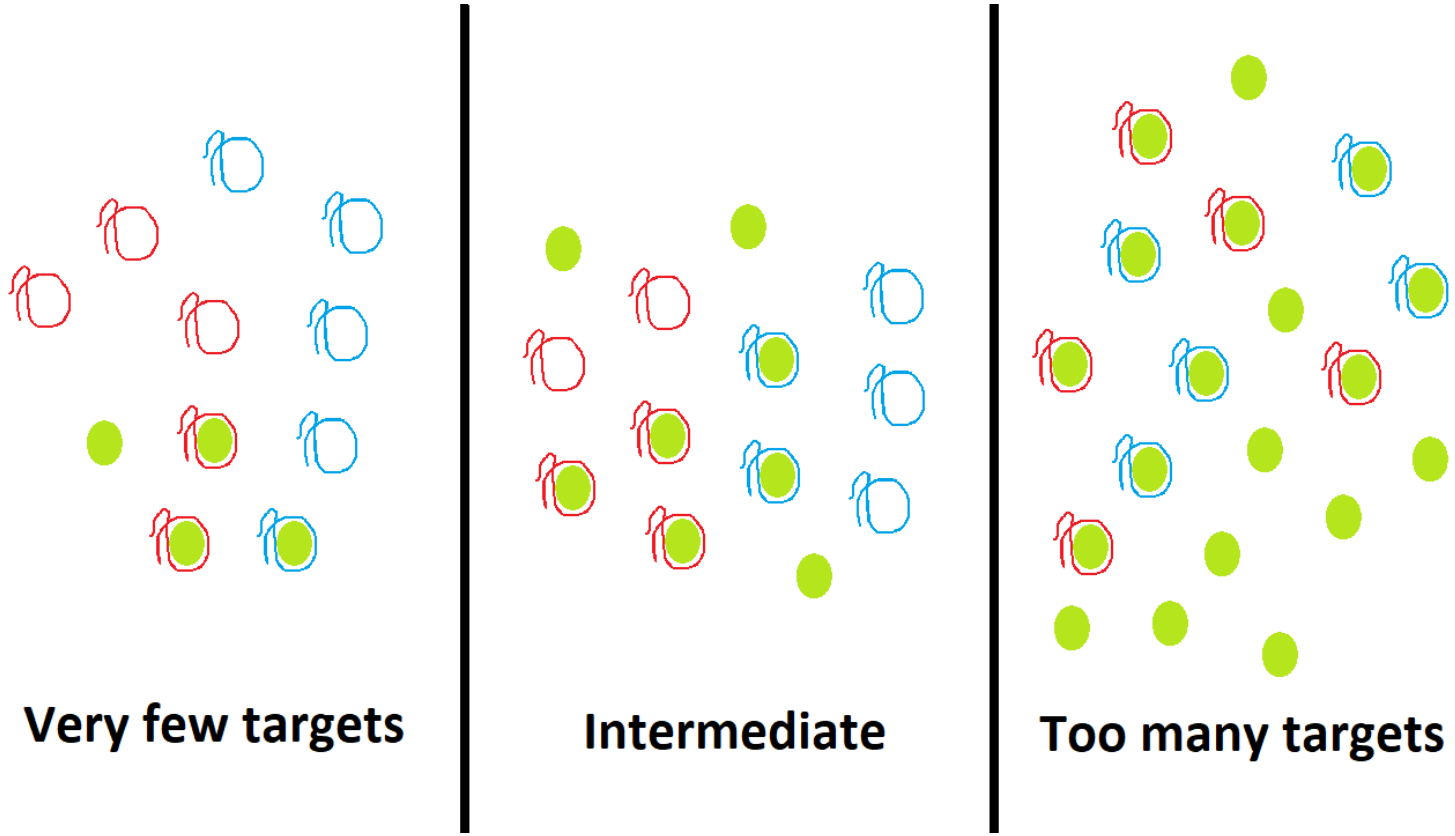
unbound target | unbound aptamer | bound aptamer

- Given $[S], [A_1], [A_2], K_1, K_2$, we can solve $[a_1], [a_2]$.

Section II: Deterministic Model

- For $\mathtt{A}_1$, the stronger aptamer, the goal is to maximize its proportion in bound aptamers: $[a_1]/([a_1]+[a_2])$.
- We can set different values of target concentration $[S]$ and aptamer concentration $[A_1]+[A_2]$, but the ratio $[A_1]/[A_2]$ is fixed.
- In this deterministic model, $[a_1]/([a_1]+[a_2])$ increases with $[A_1]+[A_2]$, and decreases with $[S]$.
- The optimal policy in the deterministic model: add as many aptamers as possible, and as few targets as possible.

Optimal policy: $[A_i] \gg [S]$.



$$\frac{[a_1]}{[a_1]+[a_2]} = \frac{[A_1]K_1}{[A_1]K_1+[A_2]K_2}$$

upper bound

$$\frac{[a_1]}{[a_1]+[a_2]} = \frac{[A_1]}{[A_1]+[A_2]}$$

lower bound

Section II: Deterministic Model

- The optimal policy in the deterministic model requires very large aptamer concentration $[A_1] + [A_2]$ and very small target concentration $[S]$.
- When $[S]$ is too small, randomness is inevitable, and the law of mass action does not hold.
- We need a stochastic model.
- We will show that something is different in this stochastic model.

Section III: Stochastic Model

- Notations: S: total number of targets; A_i: total number of $\mathbb{A}_i$ type aptamers; a_i: number of $\mathbb{A}_i$ aptamers that are bound to targets. $\bar{K}_i = K_i/V$: reaction coefficient, where V is the system volume.
- Consider a continuous-time Markov chain on 2-dimensional lattice $\mathbb{Z}^2$, where the states are the bound aptamer counts (a_1, a_2).
- The transition rates satisfy

$$\frac{r[(a_1,a_2)\to(a_1+1,a_2)]}{r[(a_1+1,a_2)\to(a_1,a_2)]} = \frac{(S-a_1-a_2)(A_1-a_1)}{a_1+1}\bar{K}_1.$$

$$\frac{r[(a_1,a_2)\to(a_1,a_2+1)]}{r[(a_1,a_2+1)\to(a_1,a_2)]} = \frac{(S-a_1-a_2)(A_2-a_2)}{a_2+1}\bar{K}_2.$$

- The stationary probability distribution satisfies

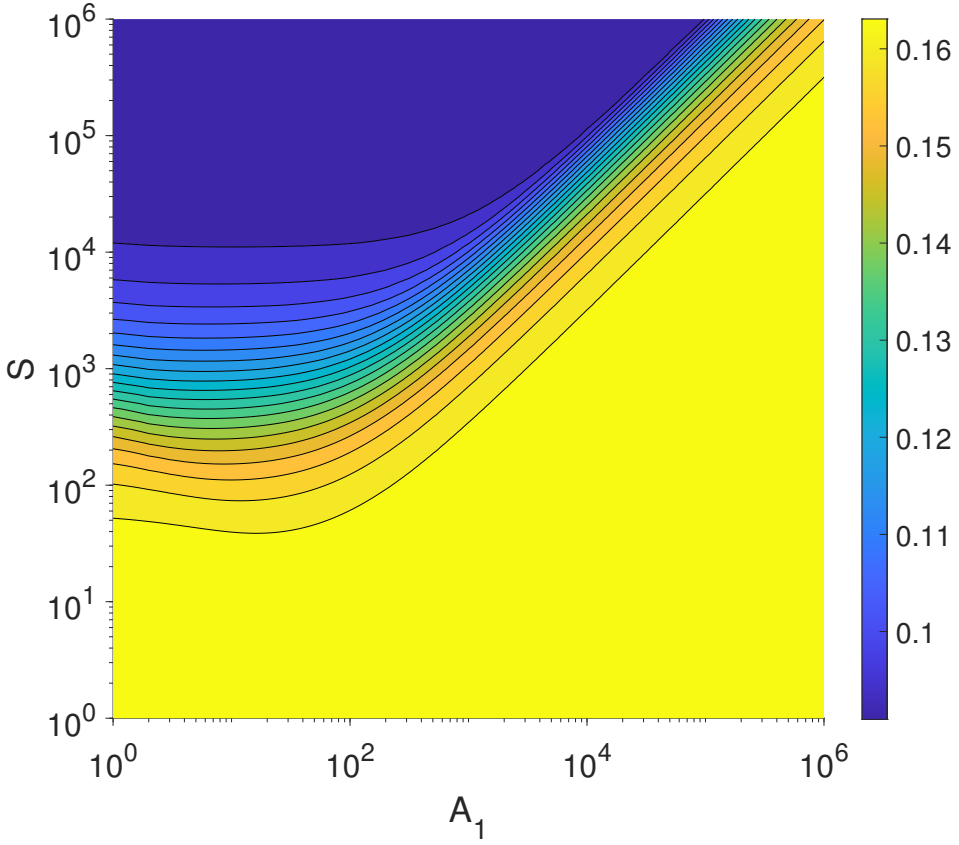
$$\mathbb{P}(a_1, a_2) = \mathbb{P}(0,0) \times \binom{S}{S - a_1 - a_2, a_1, a_2} \times \left[\binom{A_1}{a_1} \binom{A_2}{a_2} \right] \times [a_1! a_2!\,] \times \left[\bar{K}_1^{a_1} \bar{K}_2^{a_2} \right]$$

- Due to stochasticity, we need to consider the expected proportion of $\mathbb{A}_1$ and exclude the case that no aptamer is bound: $\mathbb{E}[a_1/(a_1 + a_2) \mid a_1 + a_2 > 0]$.

Section III: Stochastic Model

- In the deterministic model, $[a_1]/([a_1]+[a_2])$ increases with $[A_1]$ (fix $[A_1]/[A_2]$) and decreases with $[S]$.
- In the stochastic model, $\mathbb{E}[a_1/(a_1+a_2) \mid a_1+a_2>0]$ decreases with S, but does not always increase with A_1 (fix A_1/A_2).
- Set $S=10, \bar{K}_1=2, \bar{K}_2=1$.
  When $A_1=A_2=1$, $\mathbb{E}[a_1/(a_1+a_2) \mid a_1+a_2>0]=0.524$.
  When $A_1=A_2=2$, $\mathbb{E}[a_1/(a_1+a_2) \mid a_1+a_2>0]=0.521$.
- The problem is from the situation that $a_1=a_2=0$, which does not appear in the deterministic model.

Contour plot of the $\mathtt{A}_1$ proportion $\mathbb{E}[a_1/(a_1+a_2) \mid a_1+a_2>0]$ as a function of aptamer number A_1 (fix A_1/A_2) and target number S: it is monotonic with S, but not A_1.



In the stochastic model, we still have similar bounds for $\mathtt{A}_1$ proportion.

Theorem

$$\mathbb{E}[a_1/(a_1+a_2) \mid a_1+a_2>0] \leq A_1\bar{K}_1/(A_1\bar{K}_1+A_2\bar{K}_2).$$

Theorem

$$\mathbb{E}[a_1/(a_1+a_2) \mid a_1+a_2>0] \geq A_1/(A_1+A_2).$$

- Sketch of proof for the upper bound:
- Consider another system with $A_1\bar{K}_1/\bar{K}_2$ molecules of aptamer $\mathbb{A}'_1$ with reaction coefficient $\bar{K}_2$, A_2 molecules of aptamer $\mathbb{A}'_2$ with reaction coefficient $\bar{K}_2$, and S molecules of target.
- In this new system, two types of aptamers are the same. Due to symmetry,

$$\mathbb{E}[\frac{a'_1}{a'_1+a'_2} \mid a'_1+a'_2>0]$$
$$=\frac{A_1\bar{K}_1/\bar{K}_2}{A_1\bar{K}_1/\bar{K}_2+A_2}=\frac{A_1\bar{K}_1}{A_1\bar{K}_1+A_2\bar{K}_2}.$$

- We just need to prove $\mathbb{E}[a_1/(a_1+a_2) \mid a_1+a_2>0] \leq \mathbb{E}[a'_1/(a'_1+a'_2) \mid a'_1+a'_2>0]$.

- Sketch of proof for the lower bound:
- Consider another system with A_1 molecules of aptamer $\mathbb{A}_1''$ with reaction coefficient $\bar{K}_2$, A_2 molecules of aptamer $\mathbb{A}_2''$ with reaction coefficient $\bar{K}_2$, and S molecules of target.
- In this new system, two types of aptamers are the same. Due to symmetry,

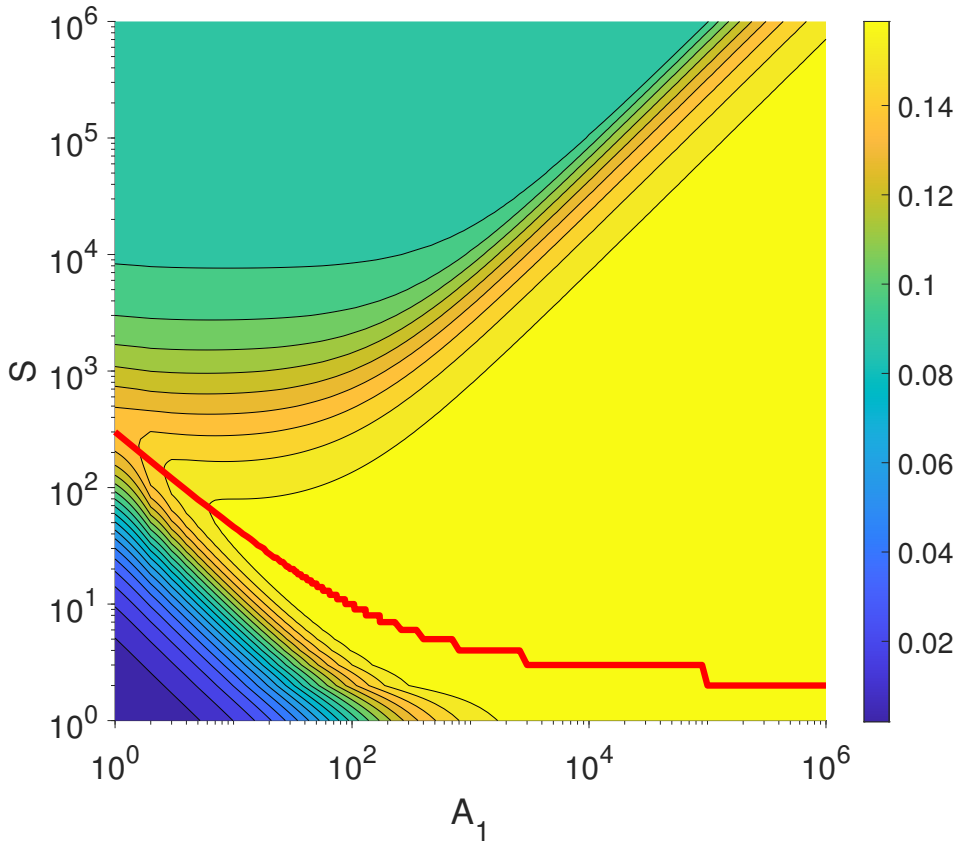
$$\mathbb{E}[\frac{a_1''}{a_1''+a_2''} \mid a_1''+a_2''>0] = \frac{A_1}{A_1+A_2}.$$

- We just need to prove $\mathbb{E}[a_1/(a_1+a_2) \mid a_1+a_2>0] \geq \mathbb{E}[a_1''/(a_1''+a_2'') \mid a_1''+a_2''>0]$.

Section III: Stochastic Model

- Studying $\mathbb{E}[a_1/(a_1+a_2) \mid a_1+a_2>0]$ is to show that the stochastic model has the same upper and lower bounds.
- In practice, we only want $\mathbb{A}_1$ aptamers. When $a_1=a_2=0$, we can stipulate that $a_1/(a_1+a_2)=0$.
- $\mathbb{E}[a_1/(a_1+a_2)] = \mathbb{E}[a_1/(a_1+a_2) \mid a_1+a_2>0] \times [1-\mathbb{P}(0,0)]$.

Section III: Stochastic Model



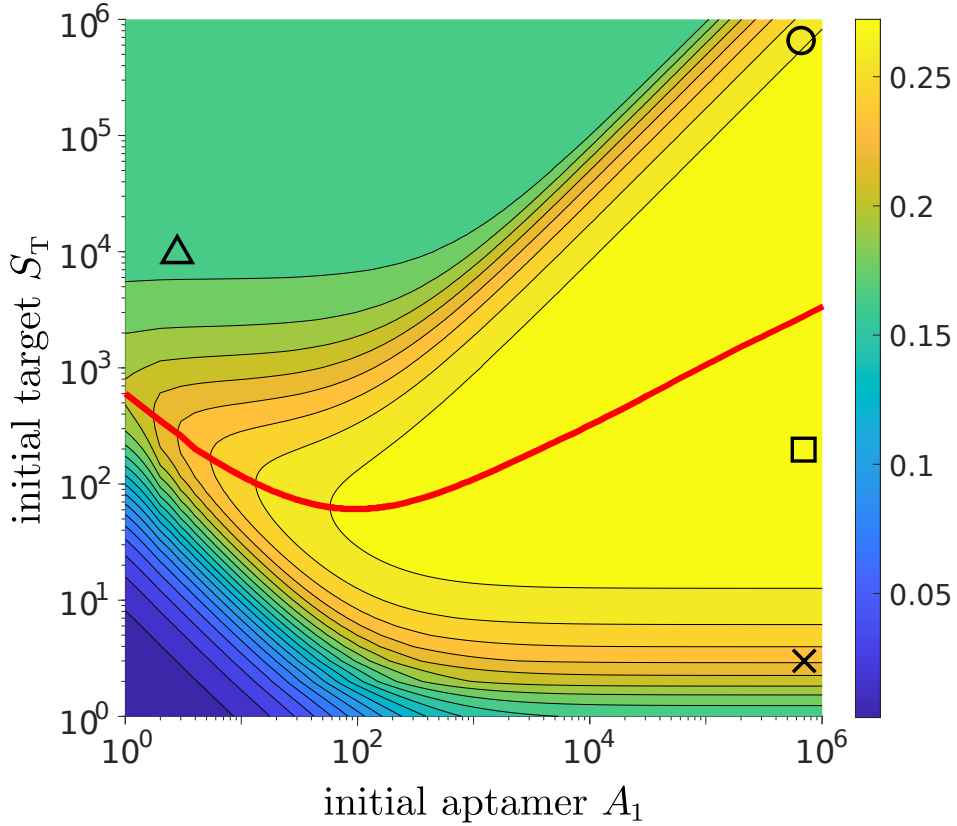
In the contour plot of $\mathbb{E}[a_1/(a_1 + a_2)]$, when the aptamer number A_1 (fix A_1/A_2) increases, the optimal target number S (red curve) decreases.

- Optimal policy in the stochastic model:
- When the aptamer number A_1 (and A_2) is not large, the target number S should not too small. Otherwise, $\mathbb{P}(a_1 = 0, a_2 = 0)$ might be large.
- What if we make A_1 (and A_2) sufficiently large, so that $\mathbb{P}(a_1 = 0, a_2 = 0) \approx 0$? Can we set $S = 1$ now?

- For one round of SELEX, $S = 1$ and very large A_1, A_2 can reach the upper bound:
  $\mathbb{E}[a_1/(a_1 + a_2)] \approx A_1\bar{K}_1/(A_1\bar{K}_1 + A_2\bar{K}_2)$.
- However, since there is only one target molecule, after one round of SELEX, only one aptamer type is left.
- After further rounds of SELEX, the expected $\mathbb{A}_1$ proportion does not increase.

Section III: Stochastic Model

- Contour plot of the $\mathbb{A}_1$ proportion $\mathbb{E}[a_1/(a_1 + a_2)]$ after two rounds of SELEX:



- For the first round, a policy with large A_1 and very small S does not perform well. The optimal target number S (red curve) first decreases and then increases with A_1.

Section III: Stochastic Model

- Theorem: Optimal policy for multiple rounds of SELEX in the stochastic model is $A_1, A_2 \gg S$ and $S \gg 1$.
- After N rounds of SELEX, this policy has $\mathbb{E}[a_1/(a_1+a_2)] \approx A_1\bar{K}_1^N/(A_1\bar{K}_1^N + A_2\bar{K}_2^N)$.
- Thus $1 - \mathbb{E}[a_1/(a_1+a_2)]$ converges to 0 exponentially fast with the rate $\approx \bar{K}_2/\bar{K}_1$. This is the most important factor for the efficiency of multi-round SELEX.

Section III: Stochastic Model

- Sketch of proof for the optimal policy:
- If the current $\mathbb{A}_1$ proportion is r, then after one round of SELEX, the expected $\mathbb{A}_1$ proportion is at most

$$f(r) = r\bar{K}_1/[r\bar{K}_1 + (1-r)\bar{K}_2],$$

  which requires $A_1, A_2 \gg S$.
- Denote the $\mathbb{A}_1$ proportion after one round of SELEX as r_1 (small S) and r_2 (sufficiently large S). We have $\mathbb{E}r_1 = \mathbb{E}r_2$ and $\text{var}(r_1) > \text{var}(r_2) \approx 0$.
- $f(r)$ is is concave (downward). By Jensen's inequality, $\mathbb{A}_1$ proportion after two rounds of SELEX satisfies $\mathbb{E}f(r_1) < \mathbb{E}f(r_2)$.
- For multiple rounds of SELEX, the optimal policy should minimize the variance of r, meaning that $S \gg 1$.

Section III: Stochastic Model

- If $\mathtt{S} + \mathtt{A}_i \rightarrow \mathtt{SA}_i$ is much easier than $\mathtt{SA}_i \rightarrow \mathtt{S} + \mathtt{A}_i$, the binding can be regarded as irreversible.
- We stop the reaction when no more binding is possible (no free aptamer or no free target).
- In this irreversible situation, we cannot have the problematic case $a_1 = a_2 = 0$.
- When $S = 1$, $\mathbb{E}[a_1/(a_1 + a_2)] = A_1\bar{K}_1/(A_1\bar{K}_1 + A_2\bar{K}_2)$; when $S \geq A_1 + A_2$, $\mathbb{E}[a_1/(a_1 + a_2)] = A_1/(A_1 + A_2)$.
- $\mathbb{E}[a_1/(a_1 + a_2)]$ is strictly decreasing with S. Thus we have the same bounds as the reversible case.
- The optimal policy is to set $S = 1$.

Section IV: Summary

- We discuss SELEX, a process to select the best aptamer for binding a target.
- In the traditional deterministic model, the optimal policy (for any rounds of SELEX) is to have a very large aptamer number and a very small target number.
- We develop a stochastic model, in which the optimal policy for multiple rounds of SELEX is to have a very large aptamer number but a moderate target number.
- This theoretical analysis can be applied to other scenarios, such as selecting drug-resistant cells.

Extra: Inheritance Law

A mathematical work in inheritance law.

Reference:
Yue Wang. (2022). “Impossibility results about inheritance and order of death.” PLOS ONE, 17(11), e0277430.

Extra: Inheritance Law

按照中国继承法，多名亲属死亡时，死亡顺序影响继承结果。

中国继承法规定了在死亡顺序无法确定时如何指定死亡顺序。

指定的死亡顺序会导致漏洞。

数学上可以证明，在满足一定基本要求的前提下，修复该漏洞有唯一办法，写在现行法国民法典中。

Extra: Inheritance Law

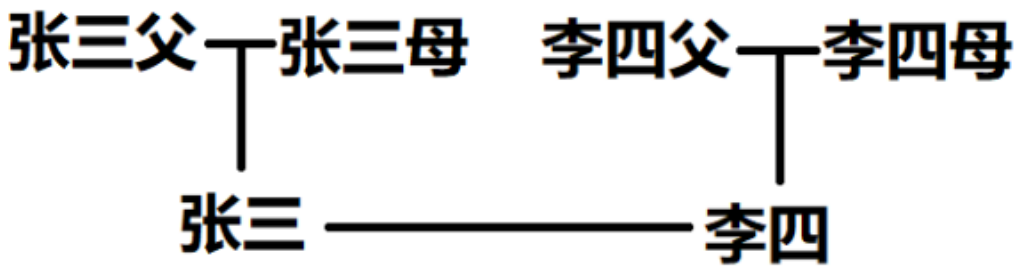
《中华人民共和国民法典》

第一千一百二十七条　遗产按照下列顺序继承：
（一）第一顺序：配偶、子女、父母；
（二）第二顺序：兄弟姐妹、祖父母、外祖父母。
继承开始后，由第一顺序继承人继承，第二顺序继承人不继承；没有第一顺序继承人继承的，由第二顺序继承人继承。

第一千一百三十条　同一顺序继承人继承遗产的份额，一般应当均等。

Extra: Inheritance Law

死亡顺序影响继承结果：
考虑一对夫妻张三、李四。他们没有孩子，但父母均在世。如果张三先于李四死亡，李四会继承张三财产的1/3，最终李四的父母继承两人财产的2/3。如果李四先于张三死亡，张三会继承李四财产的1/3，最终李四的父母继承两人财产的1/3。



Extra: Inheritance Law

如果死亡顺序未知，则继承结果无法确定。

《民法典》第一千一百二十一条　继承从被继承人死亡时开始。
相互有继承关系的数人在同一事件中死亡，难以确定死亡时间的，推定没有其他继承人的人先死亡。都有其他继承人，辈份不同的，推定长辈先死亡；辈份相同的，推定同时死亡，相互不发生继承。

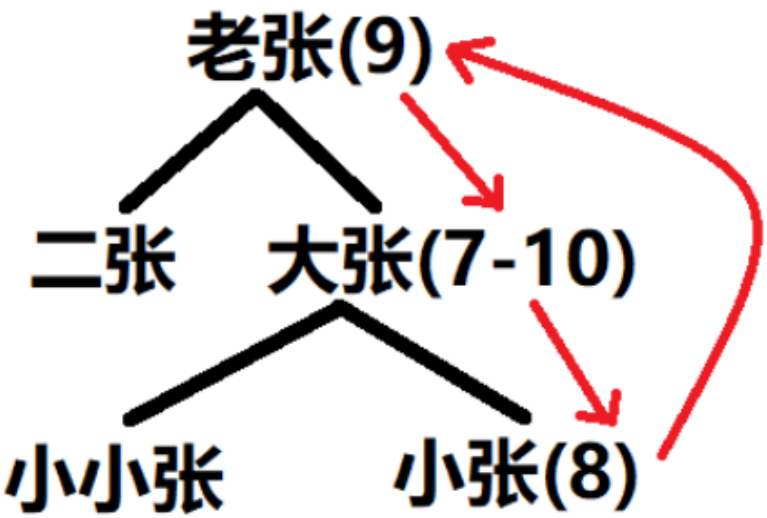
Extra: Inheritance Law

“辈分”本身并不总是能够良定义。假设王二与其舅母的妹妹结婚，那么王二、王二的舅舅、王二的舅母、王二的舅母的妹妹四人的辈分如何比较？（现实中存在的例子：秘鲁作家巴尔加斯·略萨。）

即使不考虑辈分无法定义的问题，在死亡顺序无法确定时，人为指定死亡先后仍有无法处理的问题。

Extra: Inheritance Law

中国继承法中的漏洞：
考虑老张，老张的孩子大张、二张，大张的孩子小张、小小张。他们没有其他在世的亲属。老张、大张、小张在事故中死亡。老张被发现时生命垂危，于九点死亡；小张被发现时生命垂危，于八点死亡；大张被发现时已死亡，法医推定的死亡时间介于七点至十点之间。老张、大张的死亡顺序未知，按照辈分指定为老张先于大张死亡，大张可以继承老张。大张、小张的死亡顺序未知，按照辈分指定为大张先于小张死亡，小张可以继承大张。现实中可以确定小张先于老张死亡，老张可以继承小张。
小张←大张←老张←小张，循环继承，矛盾。



Extra: Inheritance Law

矛盾的来源是多人死亡，死亡顺序部分未知，指定的死亡顺序和现实中的顺序产生矛盾。
这一反例在现实中很难发生。如果未来能够星际旅行，老张、小张在地球，大张在遥远的外星，并且死亡时间接近，那么大张与老张、小张的死亡互相不在对方的光锥中。根据相对论，大张与老张、小张的死亡顺序一定无法确定。

Extra: Inheritance Law

法国现行的继承法规定“有继承关系的两人死亡，且死亡顺序无法确定时，两人互相不继承”（《法国民法典》725-1条）。这一方法相当于认为两人的死亡时间无法比较，并不指定顺序，于是不会引起矛盾。历史上的普鲁士、奥匈帝国，以及现在的俄罗斯都采用这一方案。

定理：在一些合理的限制下，继承方法（哪些亲属可以继承，各继承多少份额）一定会受死亡顺序影响（非交换），而且人为指定死亡顺序一定会引起无法决定继承结果的矛盾。唯一能解决死亡顺序问题的就是法国现在使用的方法。

Extra: Inheritance Law

证明梗概：使用离散数学技巧，提出合理的准则并导出矛盾。

1 证明继承方法一定依赖于死亡顺序

规定以下准则：
（1）继承权只依赖于亲属关系；（2）男女平等；（3）某人死亡之时，所有子女都可以继承一部分遗产；（4）假设某人死亡之时，没有在世的配偶、后代、兄弟姐妹；如果其父母只有一方在世，则在世的父或母可以继承一部分遗产；如果其父母均在世，则父母均分所有遗产。
（如果违反2-4中任何一条，则可以构造出不依赖于死亡顺序的继承方法。）

不难构造例子使得继承结果依赖于死亡顺序。

Extra: Inheritance Law

2 证明人为指定死亡顺序一定会引起无法决定继承结果的矛盾

规定以下准则：
在指定无法确定死亡顺序的两人之间能否继承时，只考虑两人的各种属性或关系，不考虑其他人。

考虑一个死亡顺序影响继承结果的情况。如果对死亡顺序未知的任意两人指定了某种顺序，则可以仿照老张、大张、小张的例子，构造一个死亡顺序部分未知的情形，使得指定的死亡顺序导致矛盾，从而无法确定继承结果。

Extra: Inheritance Law

证明的原理：考虑一个有限集（死亡的亲属）上的全序（法律指定的死亡顺序），那么不难找到一个偏序（部分已知的现实中的死亡顺序），使得没被偏序覆盖到的这部分全序（死亡顺序未知而指定的死亡顺序）和该偏序矛盾。

为避免问题，《中华人民共和国民法典》第一千一百二十一条第二款可以考虑改为“有继承关系的两人死亡，且死亡顺序无法确定时，两人互相不继承”。

Thank you!