Mathematical Biosciences 240 (2012) 201–211

Contents lists available at SciVerse ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Theoretical consideration of selective enrichment in *in vitro* selection: Optimal concentration of target molecules

Takuyo Aita [a,*], Koichi Nishigaki [b], Yuzuru Husimi [c]

[a] Graduate School of Science and Engineering, Saitama University, 255 Shimo-okubo, Saitama 338-8570, Japan
[b] Department of Functional Materials Science, Faculty of Engineering, Saitama University, 255 Shimo-okubo, Saitama 338-8570, Japan
[c] Innovation Research Organization, Saitama University, 255 Shimo-okubo, Saitama 338-8570, Japan

ARTICLE INFO

Article history:
Received 6 January 2012
Received in revised form 20 July 2012
Accepted 21 July 2012
Available online 2 August 2012

Keywords:
SELEX
Biopanning
In vitro evolution
Selection pressure
Evolutionary design
Directed evolution

ABSTRACT

We considered an *in vitro* selection system composed of a peptide-ligand library and a single target protein receptor, and examined effective strategies to realize maximum efficiency in selection. In the system, a ligand molecule with sequence s binds to a target receptor with probability of $[R]/(K_{ds}+[R])$ (specific binding) or binds to non-target materials with probability of q (non-specific binding), where $[R]$ and K_{ds} represent the free target-receptor concentration at equilibrium and dissociation constant K_d of the ligand sequence s with the receptor, respectively. Focusing on the fittest sequence with the highest affinity (represented by $K_{d1} \equiv \min\{K_{ds}|s=1,2,\ldots,M\}$) in the ligand library with a library size N and diversity M, we examined how the target concentration $[R]$ should be set in each round to realize the maximum enrichment of the fittest sequence. In conclusion, when $N \gg M$ (that realizes a deterministic process), it is desirable to adopt $[R]=K_{d1}$, and when $N=M$ (that realizes a stochastic process), $[R]=\sqrt{K_{d1}\langle K_d^{-1}\rangle^{-1}q}$ only in the first round (where $\langle*\rangle$ represents the population average) and $[R]=K_{d1}$ in the subsequent rounds. Based on this strategy, the mole fraction of the fittest increases by $(2q)^{-r}$ times after the rth round. With realistic parameters, we calculated several quantities such as the optimal $[R]$ values and number of rounds needed. These values were quite reasonable and consistent with observations, suggesting the validity of our theory.

1. Introduction

In the last two decades, *in vitro* selection or equivalently, SELEX (systematic evolution of ligands by exponential enrichment) or biopanning has developed as one of the evolutionary drug-designs [1–4] or molecular computing technologies [5,6]. Particularly, many studies dealt with the ligand-receptor binding system as follows. A ligand library is prepared by random synthesis of peptides, antibodies, RNA or DNA molecules. The ligand library is mixed with the target molecules in a test tube, and incubated up to the equilibrium state. A part of the ligand population is bound with the target (specific binding) or non-target materials (non-specific binding). Free ligands or a part of the ligands bound with the non-target are removed by the washing process which is repeated several times. The ligands bound with the target or non-target are eluted from the mixture. The collected ligands are amplified and subject to the following selection rounds. It should be noted that the cases where random mutagenesis are introduced in the amplification process are particularly called the "*in vitro* evolution", which is comprehended as a notion of the hill-climbing on a fitness landscape in sequence space [7]. In this paper, we focus on the *in vitro* selection without the mutagenesis process.

There are two experimental parameters that govern the selection stringency: one is the concentration of the target molecules and the other is the wash time. According to the law of mass action, a ligand molecule binds to a target receptor with probability of $[R]/(K_d+[R])$ at equilibrium [8], where $[R]$ is the target concentration and K_d is dissociation constant. According to kinetics, the probability that a ligand will remain bound with the target after a wash time t is given by $\exp(-k_{off}t)$, where k_{off} is dissociation rate constant [9]. If the wash time is zero, only $[R]$ becomes the selection stringency and the ligand with the lowest K_d value becomes the winner in the selection. If the wash time is long, both $[R]$ and t become the selection stringency and the ligands with low K_d and low k_{off} become the winners [10].

Several theoretical studies on *in vitro* selection have been reported [11,9,8,12–15,10,16–18]. They examined the optimal experimental conditions such as the target concentration $[R]$ to realize the maximum enrichment of the high (or highest) affinity ligands

* Corresponding author.
E-mail address: taita@mail.saitama-u.ac.jp (T. Aita).

0025-5564/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.mbs.2012.07.006

Nomenclature

v	chain length of peptide ligands
λ	number of available letters at each site ($\lambda = 20$ for naturally occurring amino acids)
M	diversity, that is the number of different sequences in a ligand library
s and t	serial number of ligand sequences in a given library ($s, t = 1, 2, \ldots, M$)
$s = 1$	fittest sequence, that is a particular ligand sequence with the highest affinity in a ligand library
N and N'	library size, that is the number of ligand molecules in a ligand library, and the number of all the collected ligand molecules after a single selection round, respectively
n_s and n'_s	number of ligand molecules with sequence s in a ligand library, and the number of the collected ligand molecules with sequence s after a single selection round, respectively $\sum_s n_s = N$, $\sum_s n'_s = N'$
$x_s \equiv n_s/N$ and $x'_s \equiv n'_s/N'$	mole fraction of ligand molecules with sequence s before and after a single selection round
V(liter)	volume of the bulk in a test tube
N_A	Avogadro's constant ($\approx 6.02 \times 10^{23}$)
[R](mole/liter)	molar concentration of free target-receptor molecules at equilibrium state
[R]*(mole/liter)	particular [R] value that maximizes the value of x'_1 for the fittest $s = 1$
q	probability that a single ligand molecule binds with the non-target materials at equilibrium
P_s	probability that a single ligand molecule with sequence s binds with the target receptor or non-target materials at equilibrium state
K_{ds}(mole/liter) and K_{ds}^{-1}(lite/mole)	dissociation constant of the ligand with sequence s with the receptor molecules and association constant, respectively: $K_{d1} < K_{d2} < K_{d3} < \cdots < K_{dM}$
$X \equiv \ln K_d$ and $X_s \equiv \ln K_{ds}$	non-dimensional binding free energy, and that for sequence s, respectively
$\psi(X)$	probability density function of X over all the ligand sequences except the fittest $s = 1$
$\langle Q \rangle$	average of Q over all the ligand sequences except the fittest $s = 1$. $\langle Q \rangle \equiv \int_{-\infty}^{\infty} Q(X)\psi(X)\mathrm{d}X$
μ and ν	mean and variance of the Gaussian distribution
$\mathbf{E}[x]$	expectation of a variable x
$\mathbf{V}[x]$	variance of a variable x
r	number of rounds in the selective enrichment ($r = 1, 2, 3, \ldots$)

in the library. Because almost all of them were based on the law of mass action, roughly speaking, their results are consistent with one another. Irvine et al. [11], Vant-Hull et al. [14], Levitan [13] and Levine and Nilsen-Hamilton [17] referred to the optimal concentration of the target, and proposed how to schedule the target concentration throughout iterative rounds. Particularly, the former two and the last one obtained the formula that gives the optimum of [R] explicitly. In addition to the effect of association-dissociation equilibrium, [9,12,13,10] examined the effect of dissociative phases by washing. As mentioned above, as the wash time is longer, k_{off} becomes the fitness measure and the mathematical analysis becomes more complicated.

In this paper, we also refer to the effective strategy to realize the maximum enrichment of the fittest ligand in a few selection rounds, by analytical approach. In Section 2, we introduce a model of *in vitro* selection. The model is based on the law of mass action as the previous studies were, but does not include the effect of the washing operation, because mathematical analysis becomes more difficult because of the reason mentioned before. Selection dynamics should be analyzed for a deterministic process and for a stochastic process. The former is realized when there are a number of molecules for each sequence, while the latter is realized when there is a single molecule for each sequence. In Section 3, for the deterministic process, we examine the optimal target concentration to realize the maximum enrichment of the fittest ligand for a single selection round. Several formulas we obtained were similar to those obtained by them [11,14,17] because we used a model based on the law of mass action, while there were differences in details. In Section 4, we examine the effective strategy to realize the maximum enrichment of the fittest ligand through successive selection rounds, and mention, from our original viewpoint, how the target concentration [R] should be set throughout all rounds. Sections 2–4 are described concerning a simple case where the ligand and target-receptor molecules are distributed uniformly over the bulk. In real experiments of *in vitro* selection, the target-receptor molecules are frequently immobilized on the surface of matrix materials or magnetic beads. In Section 5, we extend our theory to these localized systems.

2. Model of *in vitro* selection

We consider an ensemble of ligand peptides that bind to a single target receptor in a test tube. It is possible for each ligand to bind to non-target materials in the test tube, this is known as non-specific binding. Let v and λ be the chain length of ligand peptides and the number of available letters at each site ($\lambda = 20$ for naturally occurring amino acids). We denote: L_s, free ligand molecule with sequence s; R, free target-receptor molecule; $L_s \bullet R$, molecular complex of L_s and R; U, non-target materials (such as wall of the test tube and other substances); $L_s \bullet U$, ligand L_s bound with U, respectively. For each case, $[*]$ and $[*]^\circ$ represent the numerical value of molar concentration (mole/liter) of a molecule index "$*$" at equilibrium and that in the preparation stage, respectively. $\{L_s\}$ represents a set of heterogeneous ligands over all sequences.

2.1. Protocol of in vitro selection

2.1.1. Preparation of a ligand library

A library of ligand molecules with a large variety of sequences is prepared by random synthesis. In the resulting library, the number of all the ligand molecules is denoted by N (="library size"), and the number of different sequences is denoted by M (="diversity"). For example, $N = 10^8 \sim 10^{12}$ molecules and $M = 20^5 \sim 20^{10}$ sequences in typical cases. In addition, let n_s be the number of ligand molecules with sequence s, that is $\sum_s n_s = N$, where $\sum_s$ means the sum over all the M sequences. Particularly, each peptide sequence in the initial library is generated by random synthesis. If $N \gg \lambda^v$ (= number of all possible sequences with length v), then $M \approx \lambda^v$ and $n_s \approx N/M \approx N\lambda^{-v}$, while if $N \ll \lambda^v$, then $M \approx N$ and $n_s \approx 1$.

2.1.2. Pre-screening of the ligand library

The initial ligand library is subject to a pre-screening process without target protein receptors R but with non-target materials U. Then, the library after the pre-screening does not contain the ligand molecules that show specific binding with U.