http://dx.doi.org/10.5012/bkcs.2012.33.1.31

A Strategy for Searching a Rare Event: Interpretation by Bayes' Theorem

Seung-Joon Jeon

*Department of Chemistry, Korea University, Seoul 136-713, Korea. *E-mail: sjjeon@korea.ac.kr*
Received August 11, 2011, Accepted November 2, 2011

Key Words : Molecular searching, Rare event, Facilitated targeting, Conditional probability, Bayes' theorem

There is the astonishing order in organisms of life which consist of an astronomical amount of molecules and seem to proceed their moving randomly. All known living things on earth are based on the complex webs of chemical reactions with biomolecules like proteins, DNA, RNA, and so on. Some of interactions among the complex interaction networks can be rare events stochastically because of involving only few interacting molecules in crowding of various kinds of molecules within cells.[1] One of examples is the regulation process through binding of transcription factors to DNA in a living *E. coli* cell.[2] The protein of transcript factor, called a *lac* repressor, regulates gene expression by site-specific binding to the site of chromosomal DNA, called a *lac* operon. The protein has to find a targeting site among 10^6-10^9 decoy sites on a long DNA molecule. The kinetics experiment of this process reported earlier specific binding at rate as much as 100 times faster than the diffusion limit of a bimolecular reaction by Smoluchowski relation.[3] Since then, a few facilitated mechanisms for this process were proposed to explain the discrepancy. However there has not been a clear consensus yet, even though reducing the dimensionality of diffusion process has regarded as the most probable mechanism.[4] Also many other cases for searching a rare event seem to be unclear for their mechanisms yet. So far, the mechanisms proposed for searching a rare event usually focused on a binding step of a searcher molecule to a target site among multi-steps, like a lock-and-key binding model. In this study, I propose a different view based on the probability theory called Bayes' theorem.[5] It gives an idea that an unbinding step can be more important than a binding step in searching a rear event.

Bayes' theorem shows how to determine inverse probability in conditional probability. It is also called as 'subjective probability' because it sometimes means personal belief of probability to happen a certain event. Suppose event A has occurred, given the mutually exclusive and exhaustive events B_i (i = 1, ··, j, ··, n). The probability that each event occurs is P(A) and $P(B_i)$. The conditional probability of event A given event B_j is denoted by $P(A|B_j)$, and its reverse probability of event B_j given event A, $P(B_j|A)$. Bayes' theorem is as follows:

$$P(A)\,P(B_j \mid A) = P(B_j)\,P(A \mid B_j) \tag{1}$$

This equation can be rearranged

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{P(A)} = \frac{P(B_j)P(A|B_j)}{\Sigma_{i=1}^{n} P(B_i)P(A|B_i)} \quad (i = 1, .., j, .., n \quad \Sigma_{i=1}^{n} P(B_i) = 1) \tag{2}$$

Let the theorem apply for searching a rare event, for which a molecule searches a specific binding site of a very large molecule in a cell. Assume that a searcher (a *lac* repressor protein in *E. coli* by the example mentioned above) is looking for a targeting site (a *lac* operator site of chromosomal DNA). In broad point of view, the searching proceeds as follows. At first, a searcher binds one of the decoy sites in a very large molecule. Then, the searcher keeps staying in the site or leaves for another searching. These steps undergo repeatedly until the searcher finds the right target. 'B' denotes the events of a searcher binding with a right targeting site, then the events of binding with the other wrong target sites will be 'not B'. Let 'A' denote the events that a searcher bound with any targeting site keeps binding with the site. The events that a searcher falls apart immediately after binding with any targeting site will be 'not A'. The conditional probability P(B|A), which is the probability of a really right targeting given an event that a searcher keeps binding, can be obtained by simplification of the equation (2),

$$P(B|A) = \frac{P(B)P(A|B)}{P(B)P(A|B)+P(\text{not B})P(A|\text{not B})} \quad (P(B) + P(\text{not B}) = 1) \tag{3}$$

where P(B) and P(not B) are the probability of a right targeting and a wrong targeting respectively, P(A|B) is the conditional probability of events kept binding given a right targeting, P(A| not B) is the conditional probability of events kept binding given a wrong targeting. As possibility for a right targeting increases, P(B|A) becomes close to 1, which is the maximum value for a right targeting at all times. To obtain the condition for maximum P(B|A), it is convenient that the notations of the probabilities in the equation (3) simplify as P(B|A) = y, P(A|B) = x, P(B) = a (accordingly, P(not B) = 1−a), P(A|not B) = b. Then

$$y = \frac{ax}{ax + (1 - a)b} \tag{4}$$

To get extremes, dy/dx = 0 gives the maximizing condition a(1−a)b = 0, which means a = 0 or 1, or b = 0. However, a = 0 or 1 are meaningless because a, which is P(B), should not be 0 even though it can be a very small value close to 0 and should not be 1 in a rear event. Therefore only b = 0 can give a maximum value, which is y = 1. For a right targeting at all times, P(A|not B), the probability of events kept binding given a wrong site docking, should be zero. However, there might be a mistake happened even though with very low

probability. Therefore, the condition to maximize P(B|A) is

$$P(B)\,P(A|B) >> P(\text{not } B)P(A|\text{not } B). \quad (5)$$

In a rare event, P(not B) ≈ 1, and generally, the probability of events kept binding given a right site docking, P(A|B) ≈ 1. Then, the maximizing condition will be P(B) >> P(A|not B). A rare event means that the probability of a right targeting is very small, 1 >> P(B). In order to happen such a rare event, P(A|not B) should be even much smaller than P(B). In addition, this result is always obtained no matter what value of P(A|B), the probability of events kept binding given a right site docking, is. Therefore, the key for a right targeting is to reduce the probability of a mistake rather than increase the probability of a right targeting.

In binding of transcription factor proteins to the repressor on chromosomal DNA as mentioned above, the probability of a rear event usually shows P(B) ≈ 10^{-6} or even smaller, therefore P(not B) ≈ 1, and usually P(A|B) ≈ 1. By substitution with those values in the equation (3), we obtain

$$P(B|A) = \frac{10^{-6}}{10^{-6} + P(A|\text{not } B)} \quad (6)$$

Figure 1 shows variations of the probability P(B|A) with the probability P(A|not B) by the equation (6). P(B|A) approaches to 1 as P(A|not B) decreases, as expected. If P(A|not B) is 10^{-6}, which is same as the probability of a right targeting, the probability P(B|A) is reduced to 0.5. In order to make binding more than 99% right targeting, the probability of events kept binding given a wrong site docking should be at least two order magnitudes smaller than the probability of a right targeting.

According to this study, the strategy for a right targeting in a rare event is to be recognizing a wrong site without a mistake rather than recognizing a right site with extreme accuracy. A lot of studies for searching a rare event like targeting of complex biomolecules or protein folding have usually concentrated on looking at the state of the global energy minimum. The studies for mechanisms focused on the docking state of a right targeting like the structure of binding complex, binding energy, and so on. One of models studied might be similar to the funneled energy landscape model, which is usually applied for protein folding problems.[6] A global minimum state for binding complex is searched through a rugged slope of a funnel-shaped potential or free energy surface. However, the model based on Bayes' theorem in this study proposes a different view which should be focusing on the state of a wrong targeting as well as a right targeting. For fast searching a right target in a rare event, a searcher molecule has to be detached from the site as soon as it recognizes a wrong site. It means that we should study unbinding processes more, as well as binding processes. It is hard to imagine what kinds of interaction mechanism between a searcher and targeting sites is happening to decide whether to keep binding or not. One possibility is to take overall docking dynamics into consideration through potential energy surface, including unbinding process as well as binding process. The other speculation is that there exists a flag on the right targeting site, which distinguishes the site clearly from the other sites. The flag can be regarded as a simple marker indicated for a searcher to pass by quickly if not shown. Therefore the flag should not make delicate interactions with the searcher molecule, not happening to be indecisive battle. It might be like recognizing the store looked for with its storefront sign, even though we do not know what internal structure the store has.

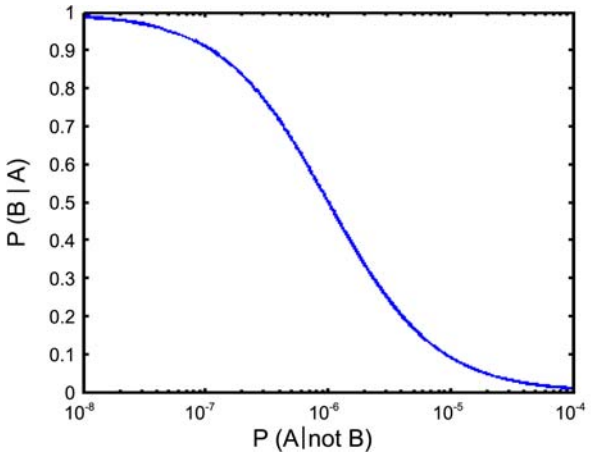

Figure 1. P(B|A) *vs.* P(A| not B) by the equation (6), in which P(B) = 10^{-6}, P(A|B) = 1, and P(not B) = 1.

Bayes' theorem gives inverse probability, which determines the posterior probability in terms of the prior probabilities estimated. There might be a philosophical disagreement regarding whether it can be used to reduce questions of objective molecular world believed in statistical behavior to problems of subjective probability like Bayes' theorem. However it has been useful to explain some of scientific problems in medicine, economics, and so on.[7] In summary, Bayes' theorem was applied to study a strategy for searching a rare event in molecular world of life. The finding is that the winning strategy is not to make a mistake, rather than to do a splendid play, like a winning tip in sports world.

References

1. Xie, X. S.; Choi, P. J.; Li, G. W.; Lee, N. K.; Lia, G. *Annual Review of Biophysics* **2008**, *37*, 417-444.
2. Tafvizi, A.; Mirny, L. A.; van Oijen, A. M. *Chemphyschem: a European Journal of Chemical Physics and Physical Chemistry* **2011**, *12*, 1481-1489.
3. Riggs, A. D.; Bourgeois, S.; Cohn, M. *J. Mol. Biol.* **1970**, *53*, 401-417.
4. von Hippel, P. H.; Berg, O. G. *J. Biol. Chem.* **1989**, *264*, 675-678.
5. Ibe, O. C. *Fundamental of Applied Probability and Random Processes*; Academic Press: Burlington, MA, USA, 2005; p 16.
6. (a) Wales, D. J. *Energy Landscapes with Applications to Clusters, Biomolecules and Glasses*; Cambridge University Press: Cambridge, UK, 2003. (b) Onuchic, J. N.; Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70-75.
7. Winnkler, R. L. *An Introduction to Bayesian Inference and Decision,* 2nd ed.; Probabilistic Publishing: Gainesville, FL, USA., 2003.