Nano Communication Networks 11 (2017) 36–41

Contents lists available at ScienceDirect

Nano Communication Networks

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

Information theoretic approach in molecular interactions and implications in molecular evolution

Y. Adiguzel

Biophysics Department, School of Medicine, Istanbul Kemerburgaz University, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 14 February 2016
Received in revised form
20 June 2016
Accepted 12 September 2016
Available online 22 September 2016

Keywords:
Biological pattern analysis
Cellular information systems
Biomolecular recognition
Pattern recognition
Shannon's information communication theory
Cellular networks

ABSTRACT

Shannon's information communication theory tells much about the reliable message transmission and has a wide scope of application in communication technologies. The aim of this work is to present the implementation of the information communication theory in the cellular recognition or molecular interaction event models to quantify the message transmission capacities of molecular binding patterns and complexes. It has implications in molecular evolution such that the pattern formation by varied combinations of a possible number of binding events does not necessitate the presence of different types of receptor–ligand complexes. Signals through the same and different types of receptor–ligand complexes seem to have equal information amounts. This is suggesting the possible role of pattern formation in differentiation of distinct conditions, maybe even changing local concentration gradients or so, through sensing the patterns of relevance. Further, recognition of the patterns that are formed up of the same binding partners would be the step preceding the recognition of the patterns with discrete binding partners. All these considerations are valid for cellular networks, wherein the communicating cells are the sources of binding target molecules and are thus imposing diffusion dependent concentration gradients and variations in the probabilities of the binding events.

1. Introduction

Shannon's information communication theory [1–3] is commonly utilized in the field of bioinformatics for biosciences [4–6]. This theory allows one to calculate the information content of a message through the magnitude of the number of possibilities that are eliminated once a certain message is received. It enables the calculation in bits due to the preference of using the logarithm with base 2. The minimal informative condition possesses 1 bit of information. Expression in bits is not sourced by a mathematical preference but the aim is the standardization of quantification with a scalable expression of the present information. By this means, information transmission capacities of diverse systems can be compared and/or limiting factors can be evaluated. The importance and feasibility of applying or adapting information theoretic concepts to the biological world was justified by the presence of redundancy in DNA [7]. The redundancy in DNA was claimed to serve in the same manner as in the utilization of redundancy for reliable message transmission through unreliable channels. All biomolecules including proteins involve in similar message communication processes, which comprise molecular interaction and binding events. Schneider reviewed the molecular information theory in 2010, where he suggested that the efficiency of protein binding is 70% according to the application of the formula equivalent to information transmission channel capacity to molecular systems [8]. Atakan and Akan introduced an information theoretical approach for molecular communication, wherein they derived a closed-form expression for the capacity of molecular communication channel and proposed adaptive error compensation [9]. Later, they used a stochastic model of molecular reactions in biochemical systems [10]. Biological systems are based on diffusion-based communication and it is suitable for the analysis and integration of sensors within the biological systems' based nanonetworks. Accordingly, a closed-form expression for the information capacity of diffusion-based molecular communication channel was obtained by Pierobon and Akyildiz [11]. Binary digital information transmission in a diffusion-based system was then designed by Meng and coworkers, as the first work that studied the signal processing, estimation, and detection issues in the presence of Inter-Symbol Interference and reception noise [12].

There has been a broad interest in molecular communication networks. As indicated by the name, this kind of networks at the molecular level is involving communication, which differentiates them from merely networked structural organizations like those of

E-mail address: yekbun.adiguzel@kemerburgaz.edu.tr.

http://dx.doi.org/10.1016/j.nancom.2016.09.002
1878-7789/

carbon nanotube networks. The term nanonetwork was also used accordingly by Akyildiz and coworkers, wherein they introduced these networks as a new paradigm [13]. Nanonetworks can be worked readily for their facts and failures through the examples that are readily available in Nature. Such a study was carried out by Malak and Akan, which addressed the concept through three major bodily communication systems, namely the neuronal, cardiac, and hormonal communication channels, together with their failures and networking problems [14]. Alternatively, MacLennan studied morphogenesis as a model of nano communication [15] and Liu studied informatics based on cells [16].

Dressler and Akan defined nanonetworking as nanomachines' communication and sharing of information to achieve a common goal [17]. Features involving such systems have much to reveal. They are studied through a wide scope based on theoretical and experimental aspects, as well as with model systems. Cakir and Oktug introduced a network environment with unique single-stranded DNAs utilized for identification of nanoelements and with messaging capabilities and implementation of database [18]. Their aim was to establish a nanonetwork environment, which would be close to the information technology. Molecule based communications were delineated by Hiyama and co-workers as slow, stochastic, low energy consuming, medium dependent and prone to the noise in it, and occurring through chemical signals that transmit information of chemical states or processes [19]. They later specified the propagation speed in molecular communication as a few micrometers per second, and the propagation distance to be ranging from nanometers to meters [20]. Energy consumption was not specified but it was indicated as extremely low. Kuran and coworkers established an energy model of communication by diffusion in molecular communication systems [21].

As stated, information theoretic approaches are helpful in information quantification and enable one to compare the amounts of information in different messages. Alternatively, information processing capacities of dissimilar mechanisms can be related. They are applied in the calculations of the information contents of molecular structures [22–25]. Here, the information communication theory is aimed to be utilized for calculating the information amounts of signals received by signaling complexes, molecular interactions, and the patterns that are formed within such types of bindings and multiplexes. Biological relevance and evolutionary significance of the resulting observations is discussed. Afterwards, the concept is broadened to networks, which will be exemplified through a simplified, two communicating cell model. Broadly, implementation of an information theoretic approach in the cellular recognition or molecular interaction events is aimed here with a critical perspective on the evaluation of their biological meanings.

2. Materials and methods

Shannon's entropy [26] is shown in Eq. (1) for a molecular interaction or one-to-one binding event.

$$H = -\sum_{i=1}^{m} P_i \log_2 P_i. \tag{1}$$

In the equation given above, the result is calculated in bits. The variable i is the binding status of the binding or interaction position. P_i is the probability of being found at the bound or unbound state. Consequently, there will always be an unbound state together with its own probability as well. In the simplest case, the possible binding status in a bimolecular interaction is the conditions of the bound and the unbound states that have equal probabilities. Therefore, P_l is 0.5 there for both conditions of being at bound or unbound state when there is only one molecule that could bind to a specific site and the probability of that site's being occupied by the target molecule is equal to that of its being unoccupied. Accordingly, m is minimum two, and it increases as the number of variable bound states for the binding position increases. Here it is assumed that there is no biased binding or interaction. So, when there is only one bound state, m in Eq. (1) would be 2 due to the presence of an unbound state as the alternative condition. In accordance, m would be 3 if there are two possible distinct bound states, m would be 4 if there are three possible distinct bound states, and so on so forth. Eq. (1) is valid for a single molecular interaction or one-to-one binding event but several interactions work together and take place concomitantly in a cell, which can be termed as signaling or binding complexes. Each of these binding events can be evaluated separately in a similar fashion as described above. Subsequently, the sum of their calculation results would be the information amount of receiving signal through that specific binding pattern. The molecular binding complexes in specific binding statuses are termed as patterns here. The information amount of signaling through such patterns would be the sum of the results of the above equations for each binding act. This sum can also be shown as in Eq. (2) when there are 3 possible bindings and all the relevant contact positions of the molecular interaction complex (*or signaling complex, which consists of signal transfer process, namely the relay of signal from the environment to the intracellular medium*) are accounted for. Please note that the complete binding reputes in the molecular interaction complex is termed here as patterns and the final H is the sum of H, calculated for each position, as follows.

$$H = \left(\underbrace{\left(-\sum_{i=1}^{m} P_i \log_2 P_i \right)}_{\text{Binding position no 1}} + \underbrace{\left(-\sum_{j=1}^{n} P_j \log_2 P_j \right)}_{\text{Binding position no 2}} + \underbrace{\left(-\sum_{k=1}^{l} P_k \log_2 P_k \right)}_{\text{Binding position no 3}} + \cdots \right). \tag{2}$$

The result of Eq. (2) is in bits. The variables i, j, and k are the binding statuses of each binding or interaction position. P_i, P_j, *and* P_k are the probabilities of separate binding events, with each having the possibility of being found at the bound or unbound states and at separate locations when there is only one-to-one binding at those places. So, there is again always an unbound state together with its own probability, for each binding position. Therefore, m, n, and l are minimum two, and increase as the number of variable bound states for each binding position increases. No incidence of biased condition is accounted for here. The total number of available positions of the multiple interaction unit in question will only be 3 in the forthcoming analysis but there can be many more, and actually there are. So, three binding positions were accounted for here and this is illustrated in Fig. 1, together with the given equations for the calculations, which results in H as the sum of H_1, H_2, and H_3, calculated for each position, as follows.

$$H = \left(\underbrace{\left(-\sum_{i=1}^{2} P_i \log_2 P_i \right)}_{\text{Binding position no 1}} + \underbrace{\left(-\sum_{j=1}^{2} P_j \log_2 P_j \right)}_{\text{Binding position no 2}} \right.$$