Information capacity of specific interactions

M. H. Huntley, A. Murugan, M. P. Brenner, PNAS **113**, 58415846 (2016).

Recommended with a commentary by Andrea J. Liu, Department of Physics and Astronomy, University of Pennsylvania

We are taught from a young age that specific interactions are the basis of biological function. Specific interactions between bases in nucleic acids preserve the integrity of the genetic code, specific interactions between proteins drive self-assembly into the machines and structures responsible for cellular function, specific interactions between receptors and antigens are the basis of immunity, and so on. It has long been recognized that both chemical specificity and physical shape complementarity (in which a protrusion in the shape of one species fits a cavity in another), are important to specificity. Until now, however, it has not been clear how to compare the relative ability of these two properties to enable specific binding. To compare chemical binding to shape complementarity on equal footing, Huntley, et al. first consider another question: if there are a large number of complementary "lock" and "key" pairs that can specifically bind to each other, how much do they end up binding to their desired partners instead of binding indiscriminately to others (crosstalk)? Clearly, the more distinct pairs, the more crosstalk. Thus, crosstalk limits the number of distinct pairs that can effectively bind to each other.

But how does one determine when crosstalk starts to overwhelm the desired binding? The key idea is that specific binding contains information, and that this information can be calculated. The authors use the concept of mutual information, as follows. Consider N species of lock and key pairs, and let binding between a lock x_i and a key y_j occur with probability $p(x_i, y_j)$, where i and j index lock and key pairs. (In a thermal system in which the binding energy of the pair is E_{ij} and all lock and key species have equal concentrations, $p(x_i, y_j) = e^{-\beta E_{ij}}/Z$.) The mutual information $I(X, Y)$ transmitted through binding is simply

$$I(X,Y) = \sum p(x_i, y_j) \log_2 \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \tag{1}$$

where $p(x_i)$ is the probability of finding x_i in a bound pair. Clearly, for small N, $I(X, Y)$ will increase with N, but once it reaches some value N_c it will no longer increase due to

cross talk. The value of $I(X, Y)$ at N_c is the "capacity" C of the lock-key pairing mechanism and provides a measure of when crosstalk starts to become significant. The great advantage of the capacity is that it provides an unbiased way to compare different specific binding mechanisms. Huntley, et al. calculate C for a simple model of chemical binding and a simple model of shape complementarity. (Spoiler alert: shape complementarity works better; in other words, it has a higher value of C.) They also show that if chemical binding is combined with shape complementarity, C is greater than the sum of the individual capacities for the two mechanisms.

Huntley, et al. apply this analysis to Pacman-shaped "lock" colloids and their partner keys, and note applications not only to self-assembling colloids with designed shapes and chemical specificities but also to proteins as signal transducers, enzymes, regulators, etc., and to the adaptive immune system, which relies on specific binding. More generally, all functional materials, from biological proteins to designed metamaterials, contain information associated with their functions. Figuring out how to quantify that information is an interesting challenge.