over the conditioning feature X_1, i.e.

$$H(X_2|X_1) = \sum_{i_1} \mathbb{P}(X_1 = x^{(1)}_{i_1}) H(X_2|X_1 = x^{(1)}_{i_1}).$$

Proposition 1 (Chain rule,).

$$H(X_1, X_2) = H(X_1) + H(X_2|X_1). \tag{1}$$

Proposition 2. *Let* $R(\Omega, X_1, X_2)$ *denote the* entropy reduction *of* Ω *first by the feature* X_1 *and then by the feature* X_2*, and* $R(\Omega, (X_1, X_2))$ *denote the* entropy reduction *of* Ω *by a pair of features* (X_1, X_2)*. Then*

$$R(\Omega, X_1, X_2) = R(\Omega, (X_1, X_2)). \tag{2}$$

The maximum entropy of an arbitrary feature is achieved when all its outcomes occur with equal probability, and this maximum value is proportional to the logarithm of the number of possible outcomes to the base 2. Thus Proposition implies that the more possible outcomes a feature has, the higher entropy reduction it could possibly lead to. Meanwhile, a feature with an arbitrary number of outcomes can be viewed as a combination of *binary features*, the ones with two possible outcomes. Even though the entropy of the combination of two features is greater than each of them, Proposition shows that partitioning the space subsequently by two features has the same entropy reduction as partitioning by their combination. Therefore, instead of considering features with outcomes as many as possible, we focus on binary features.

1. Query repeats

Here we assess the improvement of the error rates by repeating the same query twice. Let Y (or N) denote the event of the queried base pair existing (or not) in the target structure. Let y (or n) denote the event of the experiment confirming (or rejecting) the base pair. Let nn denote the event of two independent experiments both rejecting the base pair. Similarly, we have yy and yn. Utilizing the same sequences and structures as described in Fig., we estimate the conditional probabilities $\mathbb{P}(n|N) \approx 0.993$ and $\mathbb{P}(n|Y) \approx 0.055$. The *prior probability* $\mathbb{P}(Y)$ can be computed via the expected number l_1 of confirmed queried base pairs on the path, divided by the number of queries in each sample. Fig. displays the distribution of l_1 having mean around 5. Thus we adopt $\mathbb{P}(Y) = \mathbb{P}(N) = 0.5$. By Bayes' theorem, we calculate the *posterior*

$$\mathbb{P}(N|nn) = \frac{\mathbb{P}(nn|N)\mathbb{P}(N)}{\mathbb{P}(nn)} = \frac{\mathbb{P}(n|N)^2\mathbb{P}(N)}{\mathbb{P}(n|N)^2\mathbb{P}(N) + \mathbb{P}(n|Y)^2\mathbb{P}(Y)},$$

where $\mathbb{P}(nn) = \mathbb{P}(nn|N)\mathbb{P}(N) + \mathbb{P}(nn|Y)\mathbb{P}(Y)$. we have $\mathbb{P}(nn|N) = \mathbb{P}(n|N)^2$ and $\mathbb{P}(nn|Y) = \mathbb{P}(n|Y)^2$. Similarly, we compute $\mathbb{P}(Y|nn)$, $\mathbb{P}(Y|yy)$ and $\mathbb{P}(Y|yn)$