

presented (Chap. 6). There are other approaches to multiple alignments that admit efficient algorithms. These are based on a statistical view to multiple alignments. The simplest such approach works as follows. Given a multiple alignment T_1, \dots, T_k of length m , define its profile π as the string π_1, \dots, π_m of length m consisting of the following frequency vectors.

$$\pi_p(c) = \frac{\text{number of occurrences of } c \text{ in column } p}{k} \quad (2.13)$$

Thus, “character” p of the profile of a multiple alignment is the probability distribution of characters within column p of the multiple alignment. Being a string of vector characters, the notion of an alignment of a profile with a string S is well-defined. What has to be clarified is how such an alignment is scored. This is done as follows, for all p and characters x :

$$\begin{aligned} \sigma_{\text{profile,string}}(\pi_p, x) &= \sum_{\text{all characters } y} \pi_p(y) \sigma(y, x) \\ \sigma_{\text{profile,string}}(\pi_p, -) &= \sum_{\text{all characters } y} \pi_p(y) \sigma(y, -) \\ \sigma_{\text{profile,string}}(-, x) &= \sigma(-, x). \end{aligned} \quad (2.14)$$

STRING TO PROFILE ALIGNMENT

Given a scoring function σ , profile π , and string S , compute a maximum score string-to-profile alignment of π with S .

We can also align a profile π to another profile ρ . Scores must obviously be defined as follows.

$$\begin{aligned} \sigma_{\text{profile,profile}}(\pi_p, \rho_q) &= \sum_{\text{all characters } y, z} \pi_p(y) \rho_q(z) \sigma(y, z) \\ \sigma_{\text{profile,profile}}(\pi_p, -) &= \sum_{\text{all characters } y} \pi_p(y) \sigma(y, -) \\ \sigma_{\text{profile,profile}}(-, \rho_q) &= \sum_{\text{all characters } z} \rho_q(z) \sigma(-, z). \end{aligned} \quad (2.15)$$

PROFILE TO PROFILE ALIGNMENT

Given a scoring function σ and profiles π, ρ , compute a maximum score profile-to-profile alignment of π with ρ .

2.6.8 Hidden Markov Multiple Alignment

A profile of a multiple alignment can be seen as a statistical model that describes for each position (column of the alignment) the emission probabilities