Appendix: Approximation of the original model

Here we will describe how to approximate the full model in Figure 1A in the main paper by the simplified model in Figure 1B in the main paper, given that each feature has a fixed level. Recall that the joint probabilities of the original model in Figure 1A is given in Equation (1) in the main paper. First we focus on the emission probabilities of the feature observations, and show that the likelihood ratio of the emission versus the background equals the product of this likelihood ratio on all levels.

$$\frac{\Pr(F_j = 1|Z_1, \cdots Z_{T-1})}{\nu_j} = \prod_{t=1}^{T-1} \frac{\Pr(F_j = 1|Z_t)}{\nu_j} \tag{S1}$$

where ν_j is the likelihood given the background feature. From Equation (S1) we can naturally obtain

$$\Pr(F_j = 1|Z_1, \cdots Z_{T-1}) = \nu_j^{2-T} \prod_{t=1}^{T-1} \Pr(F_j = 1|Z_t)$$

for each feature, and it is combined as

$$\Pr(F|Z_1, Z_2, \cdots Z_{T-1}) = \left(\prod_j \nu_j^{2-T} \right) \prod_{t=1}^{T-1} \Pr(F|Z_t) \tag{S2}$$

The full emission probability for each feature, $\Pr(F_j|Z_1, Z_2, \cdots Z_{T-1})$, is defined as a noisy observation (with false positive and false negative) of the OR function over Z_t,

$$\Pr(F_j = 1|Z_1 = k_1, Z_2 = k_2, \cdots Z_{T-1} = k_{T-1}) = \begin{cases} \nu_j & \text{if } \forall t \ k_t \neq M + j \\ \nu_0 & \text{if } \exists t \ k_t = M + j \end{cases}$$

However the OR function is unnecessary because we require feature F_j to have a fixed level, so only one level can emit the corresponding emission such that $Z_t = k_t = M + j$. Now to prove Equation (S1), when one of the levels indeed emit the corresponding emission, we start from the right hand side of Equation (S1) and apply Equation (3) in the main paper,

$$\prod_{t=1}^{T-1} \frac{\Pr(F_j = 1|Z_t)}{\nu_j} = \frac{\nu_0 \nu_j^{T-2}}{\nu_j^{T-1}} = \frac{\nu_0}{\nu_j} = \frac{\Pr(F_j = 1|Z_1, \cdots Z_{T-1})}{\nu_j}$$

and reach the left hand side of Equation (S1). Similarly when none of the levels emit the corresponding emission,

$$\prod_{t=1}^{T-1} \frac{\Pr(F_j = 1|Z_t)}{\nu_j} = \frac{\nu_j^{T-1}}{\nu_j^{T-1}} = \frac{\nu_j}{\nu_j} = \frac{\Pr(F_j = 1|Z_1, Z_2, \cdots Z_{T-1})}{\nu_j}$$

Hence we have derived Equation (S1) given the requirement that each feature must have a fixed level.

The above derivation for feature likelihood term is exact, but approximation is necessary for the sequence likelihood term. Similar to feature observations, we approximate the likelihood ratio of emission probabilities for sequence by a set of motifs over the background likelihood as the product of this likelihood at each level,

$$\frac{\Pr(S|Z_1, Z_2, \cdots Z_{T-1})}{\Pr(S|\lambda_0)} \approx \prod_{t=1}^{T-1} \frac{\Pr(S|Z_t)}{\Pr(S|\lambda_0)} \tag{S3}$$

where λ_0 is the null model as in Equation (2) in the main paper. We assume that motifs are independent to each other since motif length is set to be short (either set to 4 peptides or 3 to 7 peptides) comparing to the sequence length, as is the case in most known targeting motifs. This is a common assumption (e.g. (Sinha, 2006)) and necessary for avoiding overfitting. However as we discussed in section 2.4 this assumption requires

that no motif is emitted twice in different levels, which is achieved by fixing the level of each feature. Similar to Equation (S2) we also write the sequence likelihood term as

$$\Pr(S|Z_1, Z_2, \cdots Z_{T-1}) = \Pr(S|\lambda_0)^{2-T} \prod_{t=1}^{T-1} \Pr(S|Z_t). \tag{S4}$$

By combining Equation (S2) and (S4), we show that the likelihood of the full model and the likelihood of the simplified model in Figure 1A and 1B respectively in the main paper is approximately up to a constant factor, so that optimizing the simplified model also optimizes the original model.

References

S. Sinha. On counting position weight matrix matches in a sequence, with application to discriminative motif finding. *Bioinformatics*, 22(14):e454–e463, Jul 2006. doi: 10.1093/bioinformatics/btl227. URL http://dx.doi.org/10.1093/bioinformatics/btl227.