on adjustment of the N_{mis} parameter

Pete Rigas, Lambert Lab

Wednesday, June 17

definition of the measure

Recall,

$$\mu(N, N_{\mathrm{mis}}, |X_{\mathrm{mis}}|, |\mathrm{mis}|, \lambda_{\mathrm{mis}}) = \frac{\prod_{i\sim j}\exp\left(-\frac{w_i}{(|\{\mathrm{mat}\}|+2)^3}\mathcal{J}_{ij}\sigma_i\sigma_j\right)\cdots}{1+\lambda_c e^{-\beta\epsilon_{\mathrm{PAM}}}+\cdots}$$

$$\frac{\cdots\prod_{i\not\sim j}\exp\left(-\frac{w_i}{(|\{\mathrm{mis}\}|+2)^3}\mathcal{J}_{ij}(1-\sigma_i\sigma_j)\right)}{\cdots(N-X_{\mathrm{mis}})^3\sum_{\mathrm{mis}}\lambda_{\mathrm{mis}}e^{-\lambda_{\mathrm{mis}}X_{\mathrm{mis}}+\ln(N_{\mathrm{mis}})}}\,.$$

this week's thoughts

in order to more precisely measure the magnitude of the exponential decrease in the transition probability at the position of mismatch, we will manipulate simple inequalities that will enable us to:

- more clearly understand the parameter ranges, which are dependent on the λ_{mis} and N_{mis} parameters that are responsible for a specific difference in visits of the random walk to a position along the protein genome,
- from which we will be able to study more than a single mismatch within a given sequence

towards a useful inequality

because we have previously composed measure vectors in previous week's presentation to generate transition probabilities for the random walk at a given position along the sequence, we can immediately look at the condition

$$\mu_N^i < \mu_N^{i+1} \ ,$$

to obtain a range of N_{mis} and λ_{mis} parameters for which the transition probability, at a position of mismatch will strictly decrease, by substituting in..

manipulating the inequality

the expressions for the measure at positions i and $i+1$ (as given in the first slide), which are of the form,

$$\frac{\prod_{i\in\mathrm{Mat}_i} e^{\frac{w_i}{\mathcal{N}}\mathcal{J}_{ij}\sigma_i\sigma_j}\prod_{i\in\mathrm{Mis}_i} e^{\frac{w_i}{\mathcal{N}}\mathcal{J}_{ij}(1-\sigma_i\sigma_j)}}{1+\lambda_p e^{-\beta\epsilon_p}+\frac{|\mathrm{Mat}_i|}{N}\lambda_c e^{-\beta\epsilon_c}+(N-X_{\mathrm{mis}})^3\sum_{\mathrm{mis}}\cdots}$$
$$<\frac{\prod_{i\in\mathrm{Mat}_{i+1}} e^{\frac{w_i}{\mathcal{N}}\mathcal{J}_{ij}\sigma_i\sigma_j}\prod_{i\in\mathrm{Mis}_{i+1}} e^{\frac{w_i}{\mathcal{N}}\mathcal{J}_{ij}(1-\sigma_i\sigma_j)}}{1+\lambda_p e^{-\beta\epsilon_p}+\frac{|\mathrm{Mat}_{i+1}|}{N}\lambda_c e^{-\beta\epsilon_c}+(N-X_{\mathrm{mis}})^3\sum_{\mathrm{mis}}\cdots},$$

where Mat_i and Mis_i denote the collection of bases at which matches and mismatches occur, up to the i and $i+1$ position, respectively, and $\mathcal{N}$ is an abbreviation for the normalization factor included in the first slide

rearranging terms from the inequality

to determine if there exists an admissible range of N_{mis} parameters so that the quantity μ^i_N strictly decreases with respect to spatial positions at which mismatches are encountered, we can break up our analysis of the measure μ into separate cases of base pair configurations along an arbitrary sequence, which include:

- a base pair match at i, and a base pair mismatch at $i+1$,
- base pair matches at i and $i+1$,
- and finally, a base pair mismatch at i and $i+1$

the focus of this week's presentation

purpose: We will exclusively deal with the case in which there is a mismatch at $i+1$, and a match at i. In this case, because we have identical terms from the product on both sides of the inequality which correspond to Mat_i, Mis_i, Mat_{i+1}, Mis_{i+1}, after dividing all terms from the inequality from the RHS to the LHS,

$$\frac{\prod_{i \in \mathrm{Mat}_{i+1}} e^{\frac{w_i}{\mathcal{N}} \mathcal{J}_{ij} \sigma_i \sigma_j} \prod_{i \in \mathrm{Mis}_{i+1}} e^{\frac{w_i}{\mathcal{N}} \mathcal{J}_{ij}(1-\sigma_i \sigma_j)}}{\prod_{i \in \mathrm{Mat}_i} e^{\frac{w_i}{\mathcal{N}} \mathcal{J}_{ij} \sigma_i \sigma_j} \prod_{i \in \mathrm{Mis}_i} e^{\frac{w_i}{\mathcal{N}} \mathcal{J}_{ij}(1-\sigma_i \sigma_j)}} = e^{\frac{w_{i+1}}{\mathcal{N}} \mathcal{J}_{i+1,j+1}(1-\sigma_{i+1,j+1})} ,$$

in which the majority of the terms, with the exception of the term for the additional mismatch at $i+1$, cancel

dealing with the remaining terms in the inequality

on the other side of the inequality,

$$\frac{1 + \lambda_p e^{-\beta\epsilon_p} + \frac{|\mathrm{Mat}_{i+1}|}{N}\lambda_c e^{-\beta\epsilon_c} + (N - X_{\mathrm{mis}})^3 \sum_{\mathrm{mis}\in\mathrm{Mis}_i}(N_{\mathrm{mis}} - X_{\mathrm{mis}})^3 \cdots}{1 + \lambda_p e^{-\beta\epsilon_p} + \frac{|\mathrm{Mat}_{i+1}|}{N}\lambda_c e^{-\beta\epsilon_c} \cdots}$$

(continued on the line below..)

$$\frac{\cdots \lambda_{\mathrm{mis}} e^{-\lambda_{\mathrm{mis}}|X_{\mathrm{mis}}-N|} + (N - X_{\mathrm{mis}_f\in\mathrm{Mis}_{i+1}})^3 e^{-\lambda_{\mathrm{mis}_{i_f}}|X_{\mathrm{mis}_{i_f}} - N|}}{\cdots + \sum_{\mathrm{mis}\in\mathrm{Mis}}(N - X_{\mathrm{mis}})^3 \lambda_{\mathrm{mis}} e^{-\lambda_{\mathrm{mis}}|X_{\mathrm{mis}}-N|}}$$

which we will further rearrange

the final expression

after further rearrangements from the whiteboard (I can post the pictures on Slack for completeness once the remaining cases have been worked out), we obtain an expression of the form,

$$\frac{1+\frac{\frac{|\mathrm{Mat}_{i+1}|}{N}\lambda_c e^{-\beta\epsilon_c}+(N-X_{\mathrm{mis}})^3 e^{-\lambda_{\mathrm{mis}_{i_f}}|X_{\mathrm{mis}_{i_f}}-N|}}{\sum_{\mathrm{mis}\in\mathrm{Mis}_{i+1}}(N-X_{\mathrm{mis}})^3\lambda_{\mathrm{mis}}e^{-\lambda_{\mathrm{mis}}|X_{\mathrm{mis}}-N|}}}{1+\frac{\frac{|\mathrm{Mat}_{i+1}|}{N}\lambda_c e^{-\beta\epsilon_c}+(N-X_{\mathrm{mis}})^3 e^{-\lambda_{\mathrm{mis}_{i_f}}|X_{\mathrm{mis}_{i_f}}-N|}}{\sum_{\mathrm{mis}\in\mathrm{Mis}_{i}}(N-X_{\mathrm{mis}})^3\lambda_{\mathrm{mis}}}} < e^{\frac{w_{i+1}}{\mathcal{N}}\mathcal{J}_{i+1,j+1}(1-\sigma_{i+1,j+1})} ,$$

after neglecting terms from the ratio which simultaneously depend on Mat_i, Mat_{i+1}, Mis_i, Mis_{i+1}

relating it all together: spatial dependence of the Hamiltonian