Posted on Authorea 19 Jan 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.161102783.33083354/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Entropy based distance cutoff for protein internal contact networks

Marcin Sobieraj[1] and Piotr Setny[2]

[1]University of Warsaw Centre of New Technologies
[2]University of Warsaw

January 19, 2021

Abstract

Protein structure networks (PSNs) have long been used to provide a coarse yet meaningful representation of protein structure, dynamics, and internal communication pathways. An important question is what criteria should be applied to construct the network so that to include relevant interresidue contacts while avoiding unnecessary connections. To address this issue we systematically considered varying residue distance cutoff length and the probability threshold for contact formation to construct PSNs based on atomistic molecular dynamics in order to assess the amount of mutual information within the resulting representations. We found that the minimum in mutual information is universally achieved at the cutoff length of 5 , irrespective of the applied contact formation probability threshold in all considered, distinct proteins. Assuming that the optimal PSNs should be characterised by the least amount of redundancy, which corresponds to the minimum in mutual information, this finding suggests an objective criterion for cutoff distance and supports the existing preference towards its customary selection around 5 length, typically based to date on heuristic criteria.

Hosted file

`Setny_main.pdf` available at https://authorea.com/users/390298/articles/504673-entropy-based-distance-cutoff-for-protein-internal-contact-networks