Soft Computing (2020) 24:603–625
https://doi.org/10.1007/s00500-019-04278-8

METHODOLOGIES AND APPLICATION

Graph coloring: a novel heuristic based on trailing path—properties, perspective and applications in structured networks

Abhirup Bandyopadhyay[1] · Amit kumar Dhar[2,3] · Sankar Basu[4,5,6]

Published online: 14 August 2019

Abstract
Graph coloring is a manifestation of graph partitioning, wherein a graph is partitioned based on the adjacency of its elements. The fact that there is no general efficient solution to this problem that may work unequivocally for all graphs opens up the realistic scope for combinatorial optimization algorithms to be invoked. The algorithmic complexity of graph coloring is non-deterministic in polynomial time and hard. To the best of our knowledge, there is no algorithm as yet that procures an exact solution of the chromatic number comprehensively for any and all graphs within the polynomial (P) time domain. Here, we present a novel heuristic, namely the 'trailing path', which returns an approximate solution of the chromatic number within P time, and with a better accuracy than most existing algorithms. The 'trailing path' algorithm is effectively a subtle combination of the search patterns of two existing heuristics (DSATUR and largest first) and operates along a trailing path of consecutively connected nodes (and thereby effectively maps to the problem of finding spanning tree(s) of the graph) during the entire course of coloring, where essentially lies both the novelty and the apt of the current approach. The study also suggests that the judicious implementation of randomness is one of the keys toward rendering an improved accuracy in such combinatorial optimization algorithms. Apart from the algorithmic attributes, essential properties of graph partitioning in random and different structured networks have also been surveyed, followed by a comparative study. The study reveals the remarkable stability and absorptive property of chromatic number across a wide array of graphs. Finally, a case study is presented to demonstrate the potential use of graph coloring in protein design—yet another hard problem in structural and evolutionary biology.

Keywords Chromatic number · Graph partitioning · NP to P · Motif identifier · Protein design

1 Introduction

In graph theory, graph coloring (Jensen and Toft 2011) is a special case of graph labeling (Díaz et al. 2002). It is an assignment of labels (Gallian 2015) traditionally known as 'colors' to edges and/or vertices of a graph subject to certain constraints. In trivial formalism, it is a way of coloring the vertices (nodes) of an undirected graph such that no two adjacent vertices could be assigned the same

Communicated by V. Loia.

Abhirup Bandyopadhyay and Sankar Basu have contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00500-019-04278-8) contains supplementary material, which is available to authorized users.

✉ Sankar Basu
nemo8130@gmail.com

[1] Department of Mathematics, National Institute of Technology, Durgapur, Mahatma Gandhi Avenue, Durgapur, West Bengal 713209, India

[2] Department of IT, IIIT Alahabad, Jhalwa, Alahabad 211012, India

[3] Present Address: Department of EECS, IIT Bhilai, Raypur 492015, India

[4] Department of Physics and Astronomy, Clemson University, Clemson, SC, USA

[5] Present Address: 3BIO, ULB, 1050 Brussels, Belgium

[6] Department of Microbiology, Asutosh College, Kolkata 700026, India

SCIENTIFIC REPORTS
nature research

OPEN

A comprehensive computational study of amino acid interactions in membrane proteins

Received: 8 May 2019
Accepted: 7 August 2019
Published online: 19 August 2019

Mame Ndew Mbaye[1,2], Qingzhen Hou[1], Sankar Basu[1], Fabian Teheux[1], Fabrizio Pucci[1,3] & Marianne Rooman[1]

Transmembrane proteins play a fundamental role in a wide series of biological processes but, despite their importance, they are less studied than globular proteins, essentially because their embedding in lipid membranes hampers their experimental characterization. In this paper, we improved our understanding of their structural stability through the development of new knowledge-based energy functions describing amino acid pair interactions that prevail in the transmembrane and extramembrane regions of membrane proteins. The comparison of these potentials and those derived from globular proteins yields an objective view of the relative strength of amino acid interactions in the different protein environments, and their role in protein stabilization. Separate potentials were also derived from α-helical and β-barrel transmembrane regions to investigate possible dissimilarities. We found that, in extramembrane regions, hydrophobic residues are less frequent but interactions between aromatic and aliphatic amino acids as well as aromatic-sulfur interactions contribute more to stability. In transmembrane regions, polar residues are less abundant but interactions between residues of equal or opposite charges or non-charged polar residues as well as anion-π interactions appear stronger. This shows indirectly the preference of the water and lipid molecules to interact with polar and hydrophobic residues, respectively. We applied these new energy functions to predict whether a residue is located in the trans- or extramembrane region, and obtained an AUC score of 83% in cross validation, which demonstrates their accuracy. As their application is, moreover, extremely fast, they are optimal instruments for membrane protein design and large-scale investigations of membrane protein stability.

Biological membranes form permeable fences between the interior of cells and the external environment. They are composed of phospholipid bilayers, which form a particular, fluid, medium that differs from the surrounding aqueous solution. A lot of proteins are embedded in, attached to, or cross the membranes. We focus here on integral membrane proteins, which cross the membrane and have thus a transmembrane, and one or two extramembrane domains.

Membrane proteins are a very important class of proteins. They play key roles in the localization and organization of the cell, as well as in the cellular function by transferring specific molecules, ions and other types of signals from the cell exterior to the interior and *vice versa*. They constitute about 30% of the entire human proteome[1]. They are the focus of a lot of pharmaceutical research, as they correspond to about 60% of the current drug targets[2].

In spite of their importance, membrane proteins have been much less studied than globular proteins. They are indeed very difficult to analyze, as their folding, native structure, stability and activity is reached only within the lipid bilayer, which complicates getting their experimental X-ray structures. Generally, their large size makes also difficult to obtain them by nuclear magnetic resonance spectroscopy. These are the reasons why transmembrane protein structures only represent about 2% of the available structures deposited in the Protein Data Bank (PDB)[3]. The analysis and modeling of the 3-dimensional (3D) structure of membrane proteins are thus key objectives for rationally guiding protein design and engineering experiments.

[1]Computational Biology and Bioinformatics, Université Libre de Bruxelles, Brussels, Belgium. [2]Department of Mathematics and Informatics, Cheikh Anta Diop University, Dakar-Fann, Senegal. [3]John von Neumann Institute for Computing, Jülich Supercomputer Centre, Forschungszentrum Jülich, Jülich, Germany. Mame Ndew Mbaye, Qingzhen Hou, Fabrizio Pucci and Marianne Rooman contributed equally. Correspondence and requests for materials should be addressed to M.R. (email: mrooman@ulb.ac.be)

BBA - Proteins and Proteomics 1868 (2020) 140474

Contents lists available at ScienceDirect

BBA - Proteins and Proteomics

journal homepage: www.elsevier.com/locate/bbapap

Criticality in the conformational phase transition among self-similar groups in intrinsically disordered proteins: Probed by salt-bridge dynamics

Abhirup Bandyopadhyay[a,1], Sankar Basu[b,*,1]

[a] Theoretical Neurosciences Group, Institute De Neurosciences Des Systems, Aix-Marseille University, France
[b] Department of Microbiology, Asutosh College (affiliated to University of Calcutta), Kolkata 700026, India

ARTICLE INFO

Keywords:
Intrinsically disordered proteins
Structural degeneracy
Self-organized criticality
Phase transition
Self-similarity and fractal geometries
Salt-bridges dynamics

ABSTRACT

Intrinsically disordered proteins (IDP) serve as one of the key components in the global proteome. In contrast to globular proteins, they harbor an enormous amount of physical flexibility enforcing them to be retained in conformational ensembles rather than stable folds. Previous studies in an aligned direction have revealed the importance of transient dynamical phenomena like that of salt-bridge formation in IDPs to support their physical flexibility and have further highlighted their functional relevance. For this characteristic flexibility, IDPs remain amenable and accessible to different ordered binding partners, supporting their potential multi-functionality. The current study further addresses this complex structure-functional interplay in IDPs using phase transition dynamics to conceptualize the underlying (avalanche type) mechanism of their being distributed across and hopping around degenerate structural states (conformational ensembles). For this purpose, extensive molecular dynamics simulations have been done and the data analyzed from a statistical physics perspective. Investigation of the plausible scope of 'self-organized criticality' (SOC) to fit into the complex dynamics of IDPs was found to be assertive, relating the conformational degeneracy of these proteins to their functional multiplicity. In accordance with the transient nature of 'salt-bridge dynamics', the study further uses it as a probe to explain the structural basis of the proposed criticality in the conformational phase transition among self-similar groups in IDPs. The analysis reveal scale-invariant self-similar fractal geometries in the structural conformations of different IDPs. The insights from the study has the potential to be extended further to benefit structural tinkering of IDPs in their functional characterization and drugging.

1. Introduction

Complex systems exist in *Nature* where the structure of the system, in an abstract sense, degenerates among population ensembles of various conformations [1]. In other words, degeneracy offers a complex system the ability and flexibility to yield the same essential output by its structurally different elements [1]. Here structure may refer to the structural conformations in quantum particles [2,3], isomerism in stereo-chemistry, rotameric variation in amino acid side-chains [4] and to wherever the concept of degenerate states in structural ensembles may be applicable. Even the synchronization pattern of electrical activities of neurons in different parts of the brain [5] or pattern of economic consumption among different social groups [6] may be mapped to an abstract structural ensemble consisting of degenerate states. Structural degeneracy is important as it provides the system the flexibility to exhibit different properties, switch between different *modus operandii*, which, in the context of living systems (or functional unit of a living system, say, proteins), supports a variety of housekeeping as well as additional functionalities. Degeneracy further imparts in a complex system a natural buffering ability, wherein, the system can readily produce distributed systemic responses to local perturbations through 'networked buffering' [7].

In spite of being integral to complex systems, the concept of degeneracy is relatively new in biology. Even, within the broad biological spectrum, it is more established in certain areas (e.g., in genetic code, immune systems etc.) [1] than others. For example, its plausibility is still to be explored in full details in highly dynamic biological soft matters and bio-polymers. Degeneracy also aids critically in evolution, wherein, the network-view of biological systems (across a wide range of spacio-temporal dimensions) often proves to be handy. It has been well argued, supported by network based theoretical analyses [8,9] that while evolution is essentially direction-less [10], degeneracy is

* Corresponding author.
E-mail address: nemo8130@gmail.com (S. Basu).
[1] Both the authors contributed equally to the work.

https://doi.org/10.1016/j.bbapap.2020.140474
Received 6 May 2020; Received in revised form 12 June 2020; Accepted 15 June 2020
Available online 21 June 2020