



TI2BioP: Topological Indices to BioPolymers

Guillermin Agüero-Chapin^{1,2*}, Reinaldo Molina-Ruiz² and Agostinho Antunes^{1,3}

- ¹ CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas, 177, 4050-123 Porto, Portugal; E-Mail: <u>gchapin@ciimar.up.pt</u>; <u>aantunes@ciimar.up.pt</u>
- ² Centro de Bioactivos Químicos, Universidad Central "Marta Abreu" de Las Villas (UCLV), Santa Clara, 54830, Cuba; E-Mail: <u>reymolina@uclv.edu.cu</u>
- ³ Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal; E-Mail: <u>aantunes@ciimar.up.pt</u>

* Author to whom correspondence should be addressed; E-Mail: <u>gchapin@ciimar.up.pt</u> *Published: 4 December 2015*

Abstract: TI2BioP (Topological Indices to BioPolymers) is a software to estimate topological indices (TIs) from two-dimensional (2D) graphical approaches for the natural biopolymers DNA, RNA and proteins. The methodology mainly turns long biopolymeric sequences into 2D artificial graphs such as Cartesian and four-color maps but also reads other 2D graphs from the thermodynamic folding of DNA/RNA strings inferred from other programs. The topology of such 2D graphs is either encoded by node or adjacency matrixes for the calculation of the spectral moments as TIs. These numerical indices were used to build up alignment-free models to the functional classification of biosequences and to calculate alignment-free distances for phylogenetic purposes. We released the version 2.0 of the software that can be freely downloaded from http://ti2biop.sourceforge.net/.

Keywords: 2D graphs; Topological indices; Alignment-free models; phylogenetics

1. TI2BioP software

TI2BioP was mainly developed from the **TOPS-MODE** methodology [1] for the estimation of the spectral moments series as TIs, but it takes advantage of the **MARCH-INSIDE** program platform [2]. It was built up on objectoriented Free Pascal IDE Tools (Lazarus) running on either a Windows or Linux operating system. TI2BioP has a friendly interface allowing users to introduce multiple fasta files containing either DNA or protein sequences to select the biopolymer 2D representation type and the calculation of TIs. We released version 2.0 of

Mol2Net, **2015**, 1(*Section F*), pages 1-3, *Proceedings* <u>http://sciforum.net/conference/mol2net-1</u>

the software that can be freely downloaded from <u>http://ti2biop.sourceforge.net/</u>. This version contains two main types of 2D artificial representations, one based on Cartesian representation for DNA strings introduced by Nandy [3] and the other inspired by the four-color maps reported by Randic [4] (**Figure 1**).

These two 2D artificial graphs implemented in **TI2BioP** can be applied to nucleotide and amino acid strings as well as to the spectral

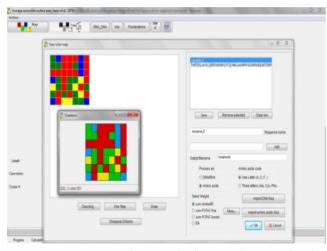


Figure 1. TI2BioP window view of the (Topological Indices to BioPolymers) software for the representation of protein four-color maps

TI2BioP can also import files containing 2D structures inferred by other DNA/RNA folding algorithms, e.g. Mfold implemented in the RNA structure software [8], for the calculation of the spectral moments as TIs. **TI2BioP** automatically represents natural biopolymers as 2D graphs and straightforward calculates spectral moments series (TIs) to be used either for statistical classification techniques in building alignmentfree models for functional classification or for deriving alignment-free several distance matrices. e.g. Euclidean, Jensen-Shannon,

moments calculations for each type of 2D DNA and protein maps [5]. It is noteworthy that the 2D Cartesian representation was extended to proteins by our group [6] and protein four-color maps were modified according to the amino acid clustering proposed in ref. [6]. Such four-color map modifications allow the speeding up of graph-building and facilitates the calculation of spectral moments as TIs [7].

Hamming and Minkowsk for phylogenetic purposes (**Figure 2**)

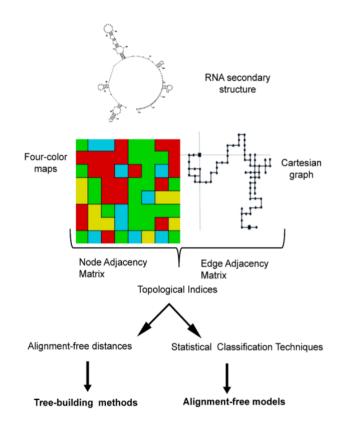


Figure 2. Workflow for the calculation of the topological indices by TI2BioP (Topological Indices to BioPolymers) from several 2D graphs for DNA, RNA and proteins

References

1. Estrada E. On the topological sub-structural molecular design (TOSS-MODE) in QSPR/QSAR and drug design research. SAR QSAR Environ Res. 2000; **11**: 55-73.

- González-Díaz H, Molina-Ruiz R, Hernandez I. MARCH-INSIDE v3.0 (MARkov CHains INvariants for SImulation & DEsign). 3.0 ed2007. p. Windows supported version under request to the main author contact email: <u>gonzalezdiazh@yahoo.es</u>.
- 3. Nandy A. Two-dimensional graphical representation of DNA sequences and intron-exon discrimination in intron-rich sequences. Comput Appl Biosci. 1996; **12**: 55-62.
- Randic M, Lers N, Plavšić D, Basak S, Balaban A. Four-color map representation of DNA or RNA sequences and their numerical characterization. Chemical Physics Letters 2005; 407: 205-8.
- 5. Molina R, Agüero-Chapin G, Pérez-González MP. TI2BioP (Topological Indices to BioPolymers) *version 2.0.*: Molecular Simulation and Drug Design (MSDD), Chemical Bioactives Center, Central University of Las Villas, Cuba; 2011
- Aguero-Chapin G, Gonzalez-Diaz H, Molina R, Varona-Santos J, Uriarte E, Gonzalez-Diaz Y. Novel 2D maps and coupling numbers for protein sequences. The first QSAR study of polygalacturonases; isolation and prediction of a novel sequence from Psidium guajava L. FEBS Lett. 2006; 580: 723-30
- 7. Aguero-Chapin G, Molina-Ruiz R, Maldonado E, de la Riva G, Sanchez-Rodriguez A, Vasconcelos V, et al. Exploring the adenylation domain repertoire of nonribosomal peptide synthetases using an ensemble of sequence-search methods. PLoS One. 2013; **8**: e65926.
- 8. Mathews DH. RNA secondary structure analysis using RNAstructure. Curr Protoc Bioinformatics. 2006; Chapter 12: Unit 12 6.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions defined by MDPI AG, the publisher of the Sciforum.net platform. Sciforum papers authors the copyright to their scholarly works. Hence, by submitting a paper to this conference, you retain the copyright, but you grant MDPI AG the non-exclusive and unrevocable license right to publish this paper online on the Sciforum.net platform. This means you can easily submit your paper to any scientific journal at a later stage and transfer the copyright to its publisher (if required by that publisher). (http://sciforum.net/about).