



Revista Cubana de Química

ISSN: 0258-5995

revcubanaquimica@cnt.uo.edu.cu

Universidad de Oriente

Cuba

Castillo-Garit, Juan A.; Marrero-Ponce, Yovani; Medina-Marrero, Ricardo; Romero-Zaldivar, Vicente; Torrens, Francisco; Castro, Eduardo A.
PROTEIN LINEAR INDICES IN BIOINFORMATICS STUDIES. 1. PREDICTION OF
PROTEIN STABILITY EFFECTS OF A COMPLETE SET OF ALANINE SUBSTITUTIONS
IN ARC REPRESSOR.

Revista Cubana de Química, vol. XVII, núm. 2, 2005, p. 202

Universidad de Oriente

Santiago de Cuba, Cuba

Available in: <http://www.redalyc.org/articulo.oa?id=443543686047>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System

Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal

Non-profit academic project, developed under the open access initiative

PROTEIN LINEAR INDICES IN BIOINFORMATICS STUDIES. 1. PREDICTION OF PROTEIN STABILITY EFFECTS OF A COMPLETE SET OF ALANINE SUBSTITUTIONS IN ARC REPRESSOR.

Juan A. Castillo-Garit,¹ Yovani Marrero-Ponce,² Ricardo Medina-Marrero,³ Vicente Romero-Zaldivar,⁴ Francisco Torrens,⁵ and Eduardo A. Castro.⁶

¹*Applied Chemistry Research Center and Department of Drug Design, Chemical Bioactive Center. Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba. Fax: 53-42-281130, 281455; Telephone: 53-42-281192, 281473, 281510; e-mail: juancg@qf.uclv.edu.cu; jacgarit@yahoo.es or juancg.22@gmail.com*

²*Department of Pharmacy, Faculty of Chemical-Pharmacy and Department of Drug Design, Chemical Bioactive Center. Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba.*

³*Department of Microbiology, Chemical Bioactive Center. Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba.*

⁴*Faculty of Informatics. University of Cienfuegos, 55500, Cienfuegos, Cuba.*

⁵*Institut Universitari de Ciència Molecular, Universitat de València, Dr. Moliner 50, E-46100 Burjassot (València), Spain.*

⁶*INIFTA, División Química Teórica, Suc.4, C.C. 16, La Plata 1900, Buenos Aires, Argentina*

A novel approach to bio-macromolecular design from a linear algebra point of view is introduced. Protein's total (whole-protein) and local (one or more amino-acid) linear indices are a new set of bio-macromolecular descriptors of relevance to protein QSAR/QSPR studies. These amino-acid level biochemical descriptors are based on the calculation of linear maps on \mathcal{R}^n [$f_k(x_{mi}): \mathcal{R}^n \rightarrow \mathcal{R}^n$] in canonical basis. These bio-macromolecular indices are calculated from the k^{th} power of the macromolecular pseudograph's α -carbon atom adjacency matrix. Total linear indices are linear functional on \mathcal{R}^n . That is, the k^{th} total linear indices are a linear maps from \mathcal{R}^n to the scalar \mathcal{R} [$f_k(x_m): \mathcal{R}^n \rightarrow \mathcal{R}$]. Thus, the k^{th} total linear indices are calculated by summing the amino-acid linear indices of all amino-acids in the protein molecule. A study of the protein stability effects for a complete set of alanine substitutions in Arc repressor illustrates this approach. A quantitative model that discriminates near wild-type stability alanine-mutants from the reduced-stability ones in a training series was obtained. This model permitted the correct classification of 97.56% (40/41) and 91.67% (11/12) of proteins in the training and test set, respectively. It show a high Matthews' correlation coefficient (MCC = 0.952) for the training set and a MCC = 0.837 for the external prediction set. Additionally, canonical regression analysis corroborated the

statistical quality of the classification model (Rcanc = 0.824). This analysis was also used to compute biological stability canonical scores for each Arc alanine-mutant. On the other hand, linear piecewise regression model compared favorably with respect to linear regression one on predicting the melting temperature (t_m) of the Arc alanine-mutants. The linear model explains almost 81% of the variance of the experimental t_m (R = 0.90 and s = 4.29) and the LOO press statistics evidenced its predictive ability (q^2 = 0.72 and s_{cv} = 4.79). Moreover, TOMOCOMD-CAMPS method produced a linear piece-wise regression (R = 0.97) between protein backbone descriptors and t_m values for alanine-mutants of Arc repressor. A break-point value of 51.87°C characterized two mutants' clusters and coincided perfectly with the experimental scale. For this reason, we can use the linear discriminant analysis and piecewise models in combination to classify and predict the stability of the mutant Arc homodimers. These models also permitted the interpretation of the driving forces of such a folding process, indicating that topologic/topographic protein's backbone interactions control the stability profile of wild-type Arc and its alanine-mutants.

Keywords: Protein Stability, Arc Repressor, Alanine-Substitution Mutant, TOMOCOMD-CAMPS Software, Protein Linear Indices, QSAR.