TARGETS OF DIMETHYL FUMARATE: A NOVEL PREDICTIVE METHOD BASED ON CONCEPTS OF NETWORK THEORY AND MACHINE LEARNING

Arianna Carolina Rosa[1], Elisa Benetti[1], Jessica Audisio[1], Gianluca Miglio[2]

[1]Department of Drug Science and Technology, University of Turin, Turin - Italy, [2]Department of Drug Science and Technology and Competence Centre for Scientific Computing C3S, University of Turin, Turin - Italy

Introduction: Dimethyl fumarate (DMF)-based drug products have been approved for use as oral treatments for psoriasis and relapsing-remitting multiple sclerosis in recent years. The adduction of DMF to certain cysteine residues in proteins is thought to underlie their effects. However, to date, only a few targets for these compounds have been discovered using proteomic methods. This study proposes a novel method by which to consistently classify (class prediction) cysteine sites in proteins in terms of their reactivity towards DMF.

Material and methods: DMF-sensitive proteins discovered in human T lymphocytes and described by Blewett et al. (Sci Signal 2016;9, rs10) were considered in this study. In particular, available 3D structure of these proteins were retrieved from the PDB repository and analysed to create the corresponding 2D representations as residue interaction networks (RIN). Cysteine sites in these RINs were first divided into two classes according to their reactivity towards DMF and then represented as numerical vectors. The vector elements were the counts of the 20 natural amino acids found in the neighbourhood (the first neighbours) of that site. The resulting vectors were combined in a $n \times 20$-matrix, were n is the total number of cysteine sites. Finally, the matrix was analysed using a panel of classification algorithms/models (Classification and Regression Tree, CART; Conditional Inference Tree, CIT; k-Nearest Neighbours, KNN; Linear Discriminant Analysis, LDA; Neuronal Network, NNET; Partial Least Square, PLS; Random Forrest, RF; Support Vector Machine, SVM).

Results: Wide-ranging and previously unforeseen variety was found in the analysis of the neighbourhood composition of cysteine sites found in DMF-sensitive proteins. Furthermore, neighbourhood composition has shown itself to be a network-type attribute that is endowed with remarkable predictive power when distinct classification algorithms were employed. The highest and lowest performances were obtained using SVM and CIT, respectively. Moreover, a comparable performance was determined for SVM, RF and NNET. These results support the adoption of neighbourhood composition as a potential predictor to estimate cysteine reactivity towards DMF.

Discussion and conclusions: The modification of cysteine sites in proteomes has led to electrophiles that are derived from DMF-based drug products exerting pleiotropic actions. The development and adoption of high-throughput and system-oriented methods and tools should be useful in interpreting the pharmacological effects of these agents. As indicated by this study, concepts of network theory, for protein structure analyses, can be combined with machine learning techniques to estimate the propensity of a cysteine residue in protein to be modified by DMF. When adopted together with other approaches for target identification/validation, this method could provide helpful data by which to find pharmacological targets of DMF-based drug products.