Targets of dimethyl fumarate: a novel predictive method based on concepts of network theory and machine learning

P.279

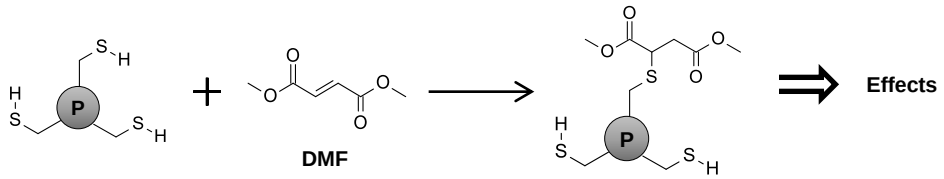
Targets of Dimethyl Fumarate: a Novel Predictive Method Based on Concepts of Network Theory and Machine Learning

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

[1]Department of Drug Science and Technology and [2]Competence Centre for Scientific Computing C[3]S, University of Turin. E-mail: gianluca.miglio@unito.it

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 (P), resulting in the appearance of the corresponding dimethyl succinated (DMSu)-P, is thought to underlie the effects of this drug.



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.

METHODS

1 Identification of DMSu-P

Literature search

32 proteins with defined 3D structure

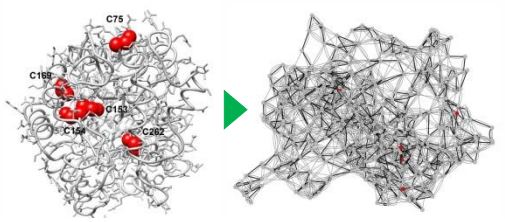
308 cysteine sites

35 modifiable cysteine (MC) sites

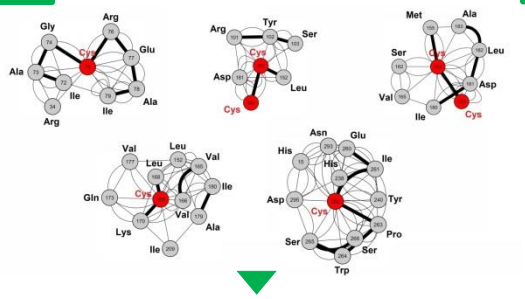
and

273 non-modifiable cysteine (NMC) sites

2 Residue Interaction Network



3 Dataset of cysteine neighbors



Neighborhood compositions

4 Data analysis

1. **Hierarchical clustering**
2. **Classification algorithms/models:**
   - **Classification and Regression Tree (CART)**
   - **Conditional Inference Tree (CIT)**
   - **k-Nearest Neighbors (KNN)**
   - **Linear Discriminant Analysis (LDA)**
   - **Neuronal Network (NNET)**
   - **Partial Least Square (PLS)**
   - **Random Forrest (RF)**
   - **Support Vector Machine (SVM)**

P.279

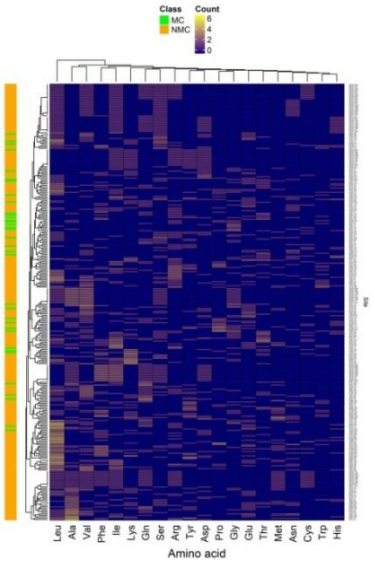
Targets of Dimethyl Fumarate: a Novel Predictive Method Based on Concepts of Network Theory and Machine Learning

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

[1]Department of Drug Science and Technology and [2]Competence Centre for Scientific Computing C[3]S, University of Turin. E-mail: gianluca.miglio@unito.it

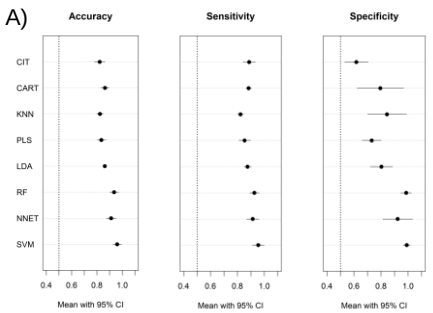
RESULTS

1 Hierarchical clustering



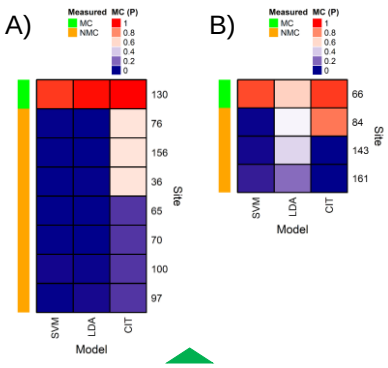
Heatmap of neighborhood compositions across all the 380 cysteine sites found in the DMF-sensitive proteins.

2 Classification



Predictive power of neighborhood amino acid composition. Eight algorithms/models (Classification and Regression Tree, CART; Conditional Inference Tree, CIT; k-Nearest Neighbors, KNN; Linear Discriminant Analysis, LDA; Neuronal Network, NNET; Partial Least Square, PLS; Random Forrest, RF; Support Vector Machine, SVM) were tested to analyze data on neighborhood compositions of cysteine site included in the DMF-DS. Accuracy, sensitivity and specificity were computed to quantify algorithm/model performance.

3 Model validation



Predicted reactivity of explicit examples. Probabilities (P) of a cysteine site found in human probable DNA dC→dU-editing enzyme (PDB ID: 3VOW; reactivity toward DMF, panel A), human tyrosyl-tRNA synthetase, cytoplasmic (1NTG; reactivity toward fumarate, panel B) to be a modifiable site was computed by analyzing data on their neighborhood using three algorithms/models and compared to the measured reactivity.

CONCLUSION

- Concepts of network theory, for protein structure analysis, 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.

REFERENCES

Blewett MM *et al.*, *Sci. Signal*. 2016, 9, rs10.

Kuhn M and Johnson K, Applied predictive modelling. Springer-Verlag, New York 2013.

Doncheva NT, *et al.*, *Nat. Protoc.* 2012, 7, 670.

This study was supported by funds from the Università degli Studi di Torino, Ricerca Locale Ex 60% 2015 and 2016-2017 to GM, and was publised in *Proteomics* 2019; 19(4):e1800301.