Accessible surface area of proteins from purely sequence information and the importance of global features[1] ESHEL FARAGGI, Dept. of Biochem. and Mol. Bio., Indiana University School of Medicine, Indianapolis, Indiana; and Research and Information Systems, LLC, Carmel, IN , YAOQI ZHOU, Institute for Glycomics and School of Informatics and Communication Technology, Griffith University, Southport Australia, ANDRZEJ KLOCZKOWSKI, Battelle Center for Mathematical Medicine, Nationwide Children's Hospital, Columbus, Ohio — We present a new approach for predicting the accessible surface area of proteins. The novelty of this approach lies in not using residue mutation profiles generated by multiple sequence alignments as descriptive inputs. Rather, sequential window information and the global monomer and dimer compositions of the chain are used. We find that much of the lost accuracy due to the elimination of evolutionary information is recouped by the use of global features. Furthermore, this new predictor produces similar results for proteins with or without sequence homologs deposited in the Protein Data Bank, and hence shows generalizability. Finally, these predictions are obtained in a small fraction (1/1000) of the time required to run mutation profile based prediction. All these factors indicate the possible usability of this work in de-novo protein structure prediction and in de-novo protein design using iterative searches.

Eshel Faraggi
IUPUI and Research and Information Systems, LLC, Carmel, Indiana

Date submitted: 14 Nov 2013   Electronic form version 1.4