Supplementary Information: Using dynamics-based comparisons to predict nucleic acid binding sites in proteins: an application to OB-fold domains

Andrea Zen[1], Cesira de Chiara[2], Annalisa Pastore[2], Cristian Micheletti[1]
[1] *SISSA, CNR-INFM Democritos and Italian Institute of Technology, Via Beirut 2, I-34014 Trieste (Italy)*
[2] *National Institute for Medical Research, The Ridgeway, London NW71AA (UK)*

I. COMPARISON BETWEEN DYNAMICS-BASED AND SEQUENCE-BASED PREDICTIONS

Several methods exist to predict nucleic acid binding sites in proteins. As an example, we are going to discuss how the dynamics-based predictions compares with the sequence-based method recently developed by Hwang *et al.* [1]. From the available on-line web server [2] we selected the mode "sequence-based binary encoding" to input the protein sequence.

The performance for the overall dataset, in terms of accuracy, sensitivity and specificity, of this method in comparison with the dynamics-based method, has already been discussed in the article. In figure 1 it is shown the dynamics-based and sequence-based predictions for OnTEBP α2, RecG and RPA70, in comparison with the actual DNA-binding residues.

[1] Hwang,S. et al. (2007) DP-Bind: a web server for sequence-based prediction of DNA-binding residues in DNA-binding proteins. Bioinformatics, **23**, 634-6.
[2] http://lcg.rit.albany.edu/dp-bind

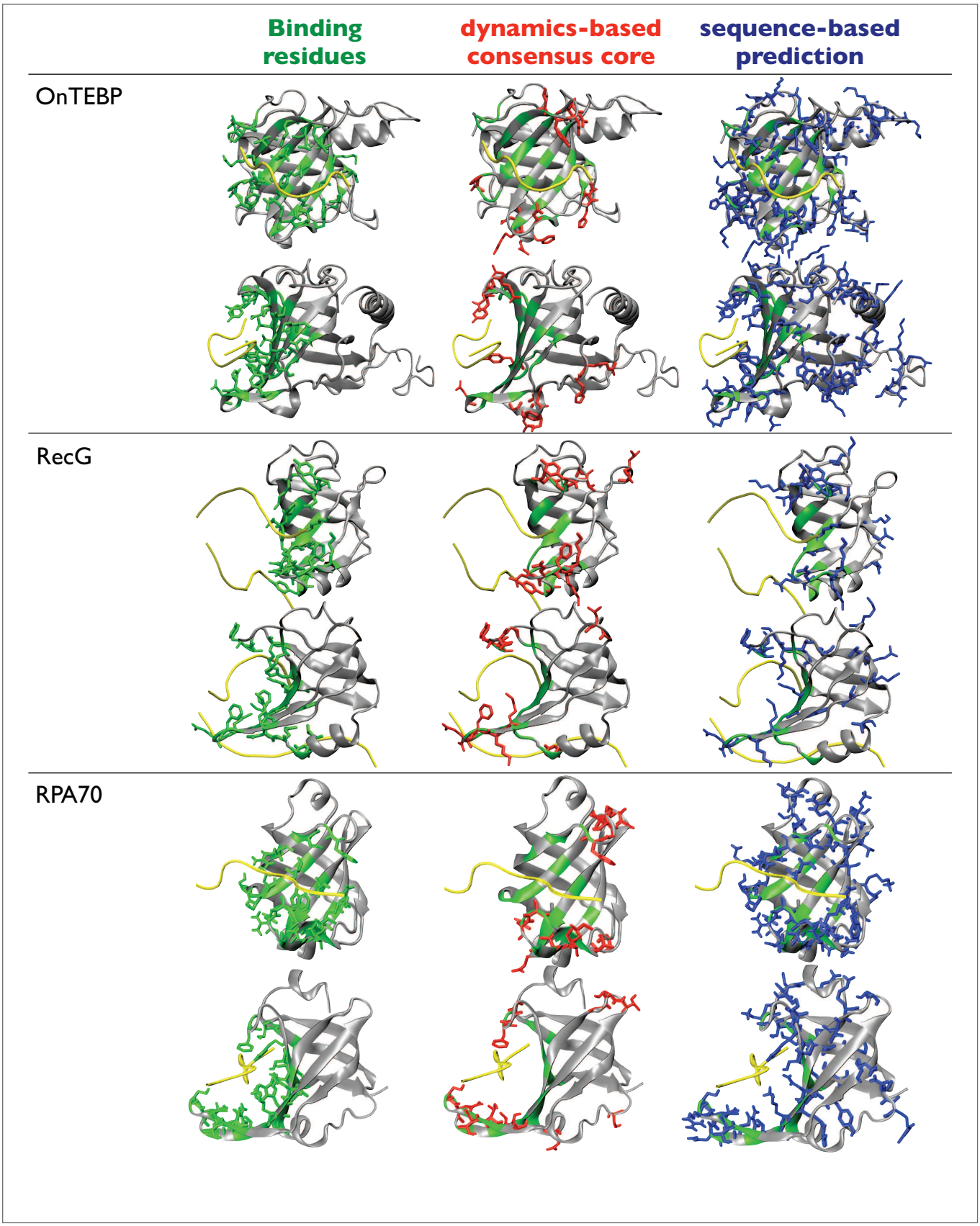

FIG. 1: Comparison between the nucleic acid binding residues, the dynamics-based consensus residues and the sequence-based DP-bind prediction[1] for DNA binding residues. Proteins here shown are OnTEBP $\alpha 2$ domain 1 (top panel), RecG (middle panel) and RPA70 repeat DBD-B (bottom panel), as in figure 3 of the article. DNA strands are shown as yellow tubes and residues that actually bind DNA are highlighted in green. Their sidechains are explicitly reported in green (first column). Sidechains highlighted in red (second column) corresponds to our dynamics-based consensus cose, and sidechains highlighted in blue (third column) correspond to a sequence-based prediction. Two different views are here displayed, as in figure 3 of the article.