

Comparison of DNA-binding across protein superfamilies

Bruno Contreras-Moreira^{1,2}, Vladimir Espinosa Angarica³

¹ Laboratory of Computational Biology, Estación Experimental de Aula Dei, Av. Montañana 1005, 50059 Zaragoza, Spain. bcontreras@eead.csic.es

² Fundación ARAID, Paseo María Agustín 36, Zaragoza, España

³ Departamento de Bioquímica y Biología Molecular y Celular, Facultad de Ciencias, Universidad de Zaragoza. Pedro Cerbuna 12, 50009 Zaragoza, España

DNA footprinting is a collection of experimental methods used to describe the sequence specificity of proteins that are known to bind DNA. The increasing collection of protein-DNA complexes reported in the Protein Data Bank encourages also computational methods that analyse sequence recognition and identify those molecular contacts that contribute most to specific binding. An example of such methods is our recently published DNAPROT algorithm [1], that predicts target sites of transcription factors based on atomistic models of protein-DNA interfaces. By applying this algorithm a collection of protein-DNA complexes was compiled and estimates of binding specificity, in the form of position weight matrices[2] and sequence logos, were obtained. Analysis of these 3D-footprints reveals that structural information is sufficient to confirm that some protein superfamilies display a higher specificity than others, as is the case of C2H2 and C2HC zinc fingers, and that different DNA-binding superfamilies show characteristic modes of recognition.

[1] Espinosa Angarica, V. *et al.* (2008) *BMC Bioinformatics*, 9:436.

[2] Lozada-Chávez, I. *et al.* (2008) *Journal of Molecular Biology*, 379(3):627-43.