Poster

Structural and target DNA-binding studies of Grainyhead/CP2 transcription factors

U. Heinemann[1], Q. Ming[1,2], Y. Roske[1], M. Rutkiewicz[1,3], J. Wang[1,4]

[1]Crystallography, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Robert-Rössle-Str. 10, 13125 Berlin, Germany. Present addresses: [2]Department of Immunology, Moffitt Cancer Center, 12902 Magnolia Dr, Tampa, FL 33612, USA,
[3]Institute of Bioorganic Chemistry, Polish Academy of Science, Noskowskiego 12/14, 61-704, Poznań, Poland,
[4]Pharmaceutical
Chemistry, Goethe-University, Max-von-Laue-Str. 15, 60438 Frankfurt am Main,

Germany heinemann@mdc-berlin.de

The family of Grainyhead/CP2 (Grh/CP2) transcription factors was originally discovered in *Drosophila* and subsequently found in a wide variety of multicellular organisms extending from insects to humans. In humans, one branch of the family is constituted by the Grainyhead-like proteins Grhl1, Grhl2 and Grhl3, and the CP2 branch comprises transcription factors Tfcp2, Tfcp2l1 and Ubp1. The CP2 sub-family members tend to be expressed ubiquitously, whereas expression of Grhl1-3 has been characterized as tissue- and developmental stage-specific. Collectively, the Grh/CP2 proteins act as transcriptional regulators of epithelial differentiation, organ development and skin barrier formation. In addition, they were assigned roles as tumor suppressors.

Since little was known about DNA target-site recognition by Grh/CP2 proteins, we started by crystallizing the DNA-binding domains of human Grhl1 and Grhl2 and by analysing their three-dimensional structures [1]. Both proteins share a common fold with the tumor suppressor p53. We also elucidated a crystal structure of target DNA-bound Grhl2, revealing the determinants of its site recognition. This structure allowed us to elucidate the molecular basis of a cancer-related single-site mutation in Grhl1.

We then extended the structural studies by genome-level studies of DNA target-site recognition by Grhl transcription factors [2, 3]. Using deep convolutional and recurrent neural networks trained with high-throughput SELEX data for Grhl1 binding, a set of non- canonical Grhl1 binding sites was identified in a data set from ChIP-Seq experiments in human cells. Surprisingly, some of binding sites thus revealed lacked the CNNG core motif found in the Grhl1-DNA co-crystal structure. These findings lead to the conclusion that a necessarily limited number of crystal structures may not suffice to cover the entire landscape of potential transcription-factor- DNA interactions.

Recently, our structural studies of gene regulation by Grh/CP2 proteins were extended by analysing structures of Tfcp2, Tfcp2l1 and a specific Tfcp2l1-DNA complex (Wang *et al*., unpublished). Although a slightly different binding mode of CP2 factors as compared to their Grh homologs had been expected, we find that the core protein-DNA interactions are closely similar in both transcription-factor sub-families.

[1] Ming, Q., Roske, Y., Schuetz, A., Walentin, K., Ibraimi, I., Schmidt-Ott, K. M. & Heinemann, U. (2018). *Nucleic Acids Res*., **46**, 2082.

[2] Leiz, J., Rutkiewicz, M., Birchmeier, C., Heinemann, U & Schmidt-Ott, K. M. (2021). *Med. Genet.*, **33**, 147.

[3] Proft, S., Leiz, J., Heinemann, U., Seelow, D., Schmidt-Ott, K. M. & Rutkiewicz, M. (2023). *BMC Genomics*., **24**, 736.