Structural basis for preferential binding of human TCF4 to DNA containing 5-carboxylcytosine

Jie Yang, John R. Horton, Xing Zhang, Xiaodong Cheng

Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

Email Contact: jieyang301@gmail.com

The psychiatric risk-associated transcription factor 4 (TCF4) is linked to schizophrenia. Rare TCF4 coding variants are found in individuals with Pitt-Hopkins syndrome – an intellectual disability and autism spectrum disorder. TCF4 contains a C-terminal basic-helix-loop-helix (bHLH) DNA binding domain which recognizes the enhancer-box (E-box) element 5'-CANNTG-3' (where N = any nucleotide). A subset of the TCF4-occupancy sites have the expanded consensus binding specificity 5'-C(A/G)-CANNTG-3', with an added outer Cp(A/G) dinucleotide such as that found in the promoter for *CNIH3*, a gene involved in opioid dependence.

In mammalian genomes, particularly brain, the CpG and CpA dinucleotides can be methylated at the 5-position of cytosine (5mC), and then may undergo successive oxidations to the 5-hydroxymethyl (5hmC), 5-formyl (5fC), and 5-carboxyl (5caC) forms. We find that, in the context of 5'-^{0}CG-^{1}CA-^{2}CG-^{3}TG-3'(where the numbers indicate successive dinucleotides), modification of the central E-box ^{2}CG has very little effect on TCF4 binding and E-box ^{1}CA modification has a negative influence on binding, while modification of the flanking ^{0}CG, particularly carboxylation, has a strong positive impact on TCF4 binding to DNA.

Crystallization of TCF4 in complex with unmodified or 5caC-modified oligonucleotides revealed that the basic region of bHLH domain adopts multiple conformations, including an extended loop going through the DNA minor groove, or the N-terminal portion of a long helix binding in the DNA major groove. The different protein conformations enable arginine 576 (R576) to interact, respectively, with a thymine in the minor groove, a phosphate group of DNA backbone, or 5caC in the major groove. The Pitt-Hopkins syndrome mutations affect five arginine residues in the basic region, two of them (R569 and R576) involved in 5caC recognition.

Our analyses indicate, and suggest a structural basis for, the preferential recognition of 5caC by this transcription factor which is centrally important in brain development.

1. Forrest, M.P., Hill, M.J., Quantock, A.J., Martin-Rendon, E. and Blake, D.J. (2014) The emerging roles of TCF4 in disease and development. Trends Mol Med, 20, 322-331.

2. Massari, M.E. and Murre, C. (2000) Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol Cell Biol, 20, 429-440.

3. Zweier, C., Peippo, M.M., Hoyer, J., Sousa, S., Bottani, A., Clayton-Smith, J., Reardon, W., Saraiva, J., Cabral, A., Gohring, I. et al. (2007) Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). Am J Hum Genet, 80, 994-1001.

4. Murre, C., Bain, G., van Dijk, M.A., Engel, I., Furnari, B.A., Massari, M.E., Matthews, J.R., Quong, M.W., Rivera, R.R. and Stuiver, M.H. (1994) Structure and function of helix-loop-helix proteins. Biochim Biophys Acta, 1218, 129-135.

5. Ren, R., Horton, J.R., Zhang, X., Blumenthal, R.M. and Cheng, X. (2018) Detecting and interpreting DNA methylation marks. Curr Opin Struct Biol, 53, 88-99.

6. Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F. et al. (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science, 356.