Nucleic Acids Research, 2023 **1**
https://doi.org/10.1093/nar/gkad583

Structural basis for specific DNA sequence motif recognition by the TFAP2 transcription factors

Ke Liu *,† **, Yuqing Xiao**†**, Linyao Gan, Weifang Li, Jin Zhang and Jinrong Min** *

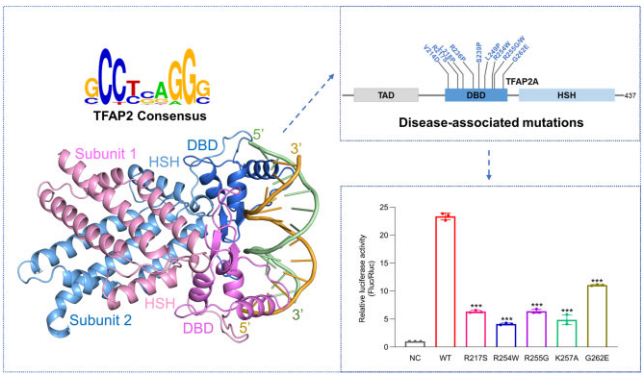
Hubei Key Laboratory of Genetic Regulation and Integrative Biology, School of Life Sciences, Central China Normal University, Wuhan 430079, PR China

Received April 24, 2023; Revised June 01, 2023; Editorial Decision June 20, 2023; Accepted June 27, 2023

ABSTRACT

The TFAP2 family regulates gene expression during differentiation, development, and organogenesis, and includes five homologs in humans. They all possess a highly conserved DNA binding domain (DBD) followed by a helix-span-helix (HSH) domain. The DBD-HSH tandem domain specifically binds to a GCC(N3)GGC consensus sequence, but the precise recognition mechanisms remain unclear. Here, we found that TFAP2 preferred binding to the GCC(N3)GGC sequence, and the pseudo-palindromic GCC and GGC motifs and the length of the central spacer between the two motifs determined their binding specificity. Structural studies revealed that the two flat amphipathic α-helical HSH domains of TFAP2A stacked with each other to form a dimer via hydrophobic interactions, while the stabilized loops from both DBD domains inserted into two neighboring major grooves of the DNA duplex to form base-specific interactions. This specific DNA binding mechanism controlled the length of the central spacer and determined the DNA sequence specificity of TFAP2. Mutations of the TFAP2 proteins are implicated in various diseases. We illustrated that reduction or disruption of the DNA binding ability of the TFAP2 proteins is the primary cause of TFAP2 mutation-associated diseases. Thus, our findings also offer valuable insights into the pathogenesis of disease-associated mutations in TFAP2 proteins.

GRAPHICAL ABSTRACT



INTRODUCTION

The TFAP2 transcription factor, initially identified as an enhancer binder of SV40, plays a crucial role in regulating gene expression and is conserved from prokaryote to eukaryotes (1). In humans, five TFAP2 homologs have been identified, known as TFAP2A-E or AP-2α, AP-2β, AP-2γ, AP-2δ and AP-2ε (2–4). These homologs are believed to have evolved from a single chordate ancestral gene, suggesting that they may share conserved functions (2). The TFAP2 proteins act as either transcriptional repressors or activators during differentiation, development, and organogenesis (5–12). In addition to their critical roles in normal biological functions, mutations and abnormal expression of the TFAP2 proteins have been linked to various diseases (13,14). Thus, the diverse functions of the TFAP2 transcription factors make them a promising area of research for both developmental biology and disease pathology.

The TFAP2 proteins contain a less conserved transactivation domain (TAD) at the N-terminus, and a highly conserved basic DNA binding domain (DBD) followed by a helix-span-helix (HSH) domain at the C-terminus (Supplementary Figure S1A). The HSH domain contributes to the dimerization of the TFAP2 proteins to form either homodimers or heterodimers, which is necessary for the

*To whom correspondence should be addressed. Email: minjinrong@ccnu.edu.cn
Correspondence may also be addressed to Ke Liu. Email: keliu2015@mail.ccnu.edu.cn
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.