⟶ **DNA ELEMENTS**

Shaping up transcription factor binding

DNA-binding proteins such as transcription factors typically display binding specificity for particular DNA sequence contexts. As the DNA sequence intrinsically influences the shape of the DNA double helix, it has been challenging to dissect the extent to which preferred binding sequences are a result of purely the underlying DNA sequence (as read-out by amino acids that make direct contact with the DNA bases) versus the steric effects of the DNA sequence in forming an ideal-shaped binding platform. A new study uses atomic-resolution structural information, high-throughput binding assays and computational modelling to quantify the relative contributions of sequence and structure effects in DNA sequences bound by Hox proteins, with implications for understanding and predicting the binding mechanisms of additional DNA-binding proteins.

Abe *et al.* built on previous work from their laboratories showing that among different *Drosophila melanogaster* Hox transcription factors, their binding specificities become more divergent when dimerized with the Extradenticle (Exd) cofactor, and that different Hox proteins in these dimers prefer DNA sequences with distinct predicted minor-groove widths. To determine whether DNA shape was making a strong contribution to specificity, they examined the known X-ray crystal structure of the Hox protein Sex combs reduced (Scr) in complex with Exd and a preferred DNA target sequence. They focused on three Scr amino acids (His–12, Arg3 and Arg5) that interact with the DNA where the minor groove is characteristically narrow but that do not directly contact the bases. Thus, these amino acids are candidates for reading the DNA structure rather than its sequence. Mutation of these residues altered the binding specificities of the Scr–Exd dimer, as determined by systematic evolution of ligands by exponential enrichment followed by high-throughput sequencing (SELEX–seq), in which a random population of short DNA oligonucleotide sequences was sequentially selected for those that bind to wild-type or mutant Scr proteins *in vitro*. Importantly, DNA structure predictions suggested that the altered sequence specificity of the mutant proteins reflected a preference for a different minor-groove width.

As further evidence of the importance of minor-groove-interacting amino acids for determining DNA shape preferences, introducing various Scr amino acids into Antennapedia (Antp) — a Hox protein with different preferences for minor-groove width — was sufficient to confer Scr–Exd binding specificity on the mutant Antp–Exd dimer, as shown by SELEX-seq *in vitro* and the ability to activate a Scr-specific reporter *in vivo*.

For a quantitative analysis of how DNA sequence and shape characteristics are predictive of binding specificities, Abe *et al.* used a machine learning approach to analyse the SELEX–seq data from various mutant and wild-type Hox proteins. The nucleotide sequence alone was a partial predictor of binding specificity, but the predictions were improved by incorporating various inferred shape features of the DNA sequences, such as minor-groove width, roll, propeller twist and helical twist. The modelling also revealed positions in the DNA sequence at which the structural features had a particularly important role in determining which Hox proteins bound. Crucially, these analyses confirmed a key role for minor-groove width at the same positions as were identified in the crystal structure, indicating that such computational modelling could be used to identify structural aspects of protein–DNA interactions even in the absence of prior structural information.

As SELEX–seq data are readily available or obtainable for various DNA-binding proteins, it will be interesting to determine the contributions of DNA shape to recognition by other proteins, and whether accounting for these spatial effects can improve our prediction and/or understanding of their genomic binding profiles.

Darren J. Burgess

ORIGINAL RESEARCH PAPER Abe, N. *et al.* Deconvolving the recognition of DNA shape from sequence. *Cell* http://www.dx.doi.org/10.1016/j.cell.2015.02.008 (2015)

PhotoDisc/Getty Images